



# Evidence of biases in genetic evaluations due to genomic preselection in dairy cattle

Clotilde C. Patry, Vincent Ducrocq

## ► To cite this version:

Clotilde C. Patry, Vincent Ducrocq. Evidence of biases in genetic evaluations due to genomic preselection in dairy cattle. *Journal of Dairy Science*, 2011, 94 (2), pp.1011-1020. 10.3168/jds.2010-3804 . hal-01000597

**HAL Id: hal-01000597**

**<https://hal.science/hal-01000597>**

Submitted on 29 May 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



## Evidence of biases in genetic evaluations due to genomic preselection in dairy cattle

C. Patry\*†<sup>1</sup> and V. Ducrocq\*

\*UMR 1313 Génétique Animale et Biologie Intégrative, Institut National de la Recherche Agronomique, 78 352 Jouy-en-Josas, France

†Union nationale des Coopératives d'Élevage et d'Insémination Animale, 149, rue de Bercy, 75 595 Paris Cédex 12, France

### ABSTRACT

A genomic preselection step of young sires is now often included in dairy cattle breeding schemes. Young sires are selected based on their genomic breeding values. They have better Mendelian sampling contribution so that the assumption of random Mendelian sampling term in genetic evaluations is clearly violated. When these sires and their progeny are evaluated using BLUP, it is feared that estimated breeding values are biased. The effect of genomic selection on genetic evaluations was studied through simulations keeping the structure of the Holstein population in France. The quality of genetic evaluations was assessed by computing bias and accuracy from the difference and correlation between true and estimated breeding values, respectively, and also the mean square error of prediction. Different levels of heritability, selection intensity, and accuracy of genomic evaluation were tested. After only one generation and whatever the scenario, breeding values of preselected young sires and their daughters were significantly underestimated and their accuracy was decreased. Genomic preselection needs to be accounted for in genetic evaluation models.

**Key words:** selection bias, genomic selection, BLUP, dairy cattle

### INTRODUCTION

In dairy cattle breeding, the recent development of genomic tools and methods has led to quick implementation of genomic selection. Due to a higher accuracy of evaluation at birth and a shorter generation interval, an increase of genetic gain is expected as well as a better management of genetic diversity. Efficiency of breeding schemes is improved, whereas their costs could be reduced. However, the use of this new strategy may damage the quality of classical genetic evaluations.

In many countries, breeding values for dairy cattle are estimated based on an animal model using BLUP methodology. Under some hypotheses, BLUP estimates have desirable properties: they are unbiased in the sense that the expected value of the prediction is equal to the expected value of what is being predicted, and they are the best among the linear predictors in the sense that they have a minimum mean squared error of prediction (**MSE**). In the mixed model equations leading to BLUP, it has been shown that the additive genetic relationship matrix (**A**), assuming an infinitesimal model, can accommodate changes in genetic means and variances due to selection (Sorensen and Kennedy, 1983). This requires that **A** is complete and correct, that the model includes all records upon which selection is based and that pedigree is complete back to the base population.

Today, genomically enhanced breeding values (**GEBV**) are computed at the birth of candidates. It allows selection of animals not only with the highest GEBV but also with the highest Mendelian sampling contribution. This selection leads to a nonrandom set of candidates being recorded for the traits of interest. Hence, the usual assumptions on Mendelian sampling expected value and variance are no longer valid. The Mendelian sampling expected value is no longer zero so that the resulting relationship matrix is no longer the correct one. Two issues are involved. The first is selective information: data on culled animals, especially records on their daughters in the case of young sires, are missing. The second is sequential selection: selection decisions are based on GEBV, and this information is not included in the evaluation model. If the genetic evaluation model remains unchanged, when records of progeny from preselected sires are included in the model, it is feared that genetic evaluations will become biased.

The effect of genomic preselection on national and international polygenic evaluations was first discussed by the Interbull Scientific Advisory Committee (Ducrocq et al., 2008) and by van der Beek (2007). Sires' ranking, genetic parameters, and genetic trend are likely to be altered at national and international levels. The conse-

Received September 9, 2010.

Accepted October 14, 2010.

<sup>1</sup>Corresponding author: clotilde.patry@jouy.inra.fr

quences of using breeding values of progeny-tested bulls on genomic prediction equations were also mentioned. Due to the rapid and widespread use of genomic evaluations worldwide (11 countries were planning to start genomic evaluations in 2009–2010; Loberg and Dürr, 2009), study of the short-term consequences on the conventional evaluation system is urgently needed.

The objective of this paper was to assess via simulation the effect of a genomic preselection step on the quality of genetic evaluations. The latter is measured by the systematic estimation error (i.e., the bias), and by the accuracy of predicted genetic effects. Comparisons were made between 2 populations, one involving a genomic preselection step of young sires [genomic preselected sires (**GPS**) population] and the other involving progeny testing only (control population). Sensitivity analysis was conducted with different levels of selection intensity, heritability, and accuracy of genomic evaluation. The effect on the genetic parameter estimates was not investigated.

## MATERIALS AND METHODS

### Overview

Bias and accuracy of genetic evaluations were measured in the GPS and control populations. The control population was expected to provide unbiased prediction and, consequently, a higher accuracy. To demonstrate bias due to genomic preselection, breeding values (true and genomically enhanced) and performance were simulated based on a real data set coming from the Holstein population in France, which included pedigree, records, and genetic parameters of the traits under consideration. Type traits were chosen for convenience (simple model, no repeated observations).

### Implementation Steps

The first key element of the chosen strategy was the identification of the candidate bulls; these consisted of the cohort of young sires (**YS**) that had only one crop of daughters in the real data set. A first BLUP evaluation based on a single trait animal model provided EBV and fixed effects estimates in the initial conditions. These EBV were used to reduce the cohort of young sires to the most promising ones on the basis of their pedigree index. Indeed, these are more likely to be genotyped by the breeding companies implementing genomic preselection. Their daughters (**D**) were then identified and their performance was deleted from the record file to mimic the data structure before a genomic preselection step. From this new data set, a second BLUP evaluation was run to obtain EBV and reliabilities (**R**), especially

of young sires' parents. The EBV and  $R^2$  were used to simultaneously simulate genomically enhanced breeding values and true breeding values (**TBV**) for the young sires and potential candidate full siblings. When genomic preselection was mimicked (GPS population), YS were selected among their full siblings, choosing the one with the highest GEBV. In both cases (with or without genomic preselection), consistent daughter records (i.e., new records) were generated, and a new BLUP evaluation was performed. As a result, EBV and TBV of the young sires and their daughters were available for bias and accuracy assessment in the 2 simulated populations. The simulation of GEBV, TBV, and daughter records, as well as the final BLUP evaluation, were repeated 50 times to measure the bias variability. Evaluations on real or simulated data were all run using an in-house BLUP software, GENEKIT, developed by the second author. The different steps of the simulation are summarized in Table 1.

### Simulation of Breeding Values and Records: Principle

Let  $a_{YS}$  be the simulated TBV of a young sire and  $\hat{a}_{YS}$  its EBV, solution of the animal mixed model equations. Let  $\hat{a}_{YS+}$  be the GEBV of a young sire combining direct genomic value and pedigree information. A selection on this GEBV mimicked the genomic preselection of young bulls before collection of phenotypic records from daughters. If **GS** and **GD** refer to the sire and dam of a young sire YS, respectively, and  $\sigma_a^2$  is the genetic variance of the trait, we have

$$\begin{aligned} a_{YS} &\sim N\left(\frac{a_{GS} + a_{GD}}{2}, \sigma_a^2\right) \\ \hat{a}_{YS} &\sim N\left(\frac{\hat{a}_{GS} + \hat{a}_{GD}}{2}, R_{YS}\sigma_a^2\right) \\ \hat{a}_{YS+} &\sim N\left(\frac{\hat{a}_{GS} + \hat{a}_{GD}}{2}, R_{YS+}\sigma_a^2\right). \end{aligned}$$

The GEBV reliability  $\left(R_{YS+}\right)$ , is a combination of the pedigree reliability  $\left(R_{PED}\right)$ , and the direct genomic reliability  $\left(R_{GEN}\right)$ , with  $R_{PED} = \frac{R_{GS} + R_{GD}}{4}$ . It was assumed that the genomic information contributes as much as  $n$  additional daughter records (Van Raden et al., 2009), so that  $R_{GEN} = \frac{n}{n+k}$ , with  $k = \frac{4}{h^2} - 1$ , and

**Table 1.** Summary of the simulation strategy and the analyses performed: description of preliminary and iterative (replicated 50 times) steps in the different populations (control, GPS)<sup>1</sup>

Step	Input data	Implementation	Output data of interest
Preliminary (1)	Data set 1: pedigree file + whole record file	BLUP 1	Initial EBV ( $\hat{a}_0$ ) + fixed effects estimates ( $\beta$ )
Preliminary (2)	Pedigree file + $\hat{a}_0$	Identification of all the YS and their D	Target cohorts: YS (size = n) and D
Preliminary (3)	Data set 2: pedigree file + record file, cohort D removed	BLUP 2	EBV and reliabilities ( $\hat{a}$ , $R$ ) of the YS parents (GS, GD): $\hat{a}_{GS}, \hat{a}_{GD}, R_{GS}, R_{GD}$
Iterative (1) control	Data set 2 + $\hat{a}_{GS}, \hat{a}_{GD}, R_{GS}, R_{GD}$	Simulation of breeding values for $n$ YS	True breeding values of the YS: $a_{YS}^C$
Iterative (2) control	Data set 2 + $\beta + \hat{a} + a_{YS}^C$	Simulation of D records	Data set 3: including the new records (true D records are replaced by the simulated ones)
Iterative (3) control	Data set 3	BLUP 3	$\hat{a}_{YS}^C, R_{YS}^C$
Iterative (4) control	$\hat{a}_{YS}^C, a_{YS}^C$	Bias assessment	$\Delta(\text{Breeding values}) = \hat{a}_{YS}^C - a_{YS}^C$
Iterative (1) GPS	Data set 2 + $\hat{a}_{GS}, \hat{a}_{GD}, R_{GS}, R_{GD}$	Simulation of breeding values for $m$ YS $\times$ $n$ GS-GD pair	GEBV + TBV of the YS: $\hat{a}_{YS}^{GPS}, a_{YS}^{GPS}$
Iterative (2) GPS	$\hat{a}_{YS+}^{GPS}$	Selection of the YS based on GEBV	Cohort of genomically selected YS
Iterative (3) GPS	Data set 2 + $\beta + \hat{a} + a_{YS}^{GPS}$	Simulation of D records	Data set 3': including the new records (true D records are replaced by the simulated ones)
Iterative (4) GPS	Data set 3'	BLUP 3'	$\hat{a}_{YS}^{GPS}, R_{YS}^{GPS}$
Iterative (5) GPS	$a_{YS}^{GPS}, \hat{a}_{YS}^{GPS}$	Bias assessment	$\Delta(\text{Breeding values})$ , comparison of reliabilities

<sup>1</sup>GPS = genomic preselection of young sires; YS = young sire; D = daughter; GS = sire of a YS; GD = dam of a YS; GEBV = genomically enhanced breeding value; TBV = true breeding value.

$h^2$  the heritability of the trait. Using Harris and Johnson's (1998) method to approximate reliability coming from different sources of information, the combined reliability ( $R_{YS+}$ ), was computed as

$$R_{YS+} = \frac{R_{PED} + R_{GEN} - 2R_{PED}R_{GEN}}{1 - R_{PED}R_{GEN}}.$$

In the same way, the genomically enhanced reliabilities  $R_{GS+}$  and  $R_{GD+}$  were computed at the GS and GD level, respectively. For example,

$$R_{GS+} = \frac{R_{GS} + \frac{1}{4}R_{GEN} - 2R_{GS}\frac{1}{4}R_{GEN}}{1 - R_{PED}R_{GEN}}.$$

It was assumed that each GS–GD pair has only one selected progeny. This is clearly not true, but this simplification avoids extra assumptions without inflating the importance of the son's GEBV on parental EBV. For young sires and their parents, all the breeding values (true, estimated, and genomically enhanced) may be defined together as coming from the following multivariate normal distribution:

$$\begin{bmatrix} a_{GS} \\ a_{GD} \\ a_{YS} \\ \hat{a}_{GS} \\ \hat{a}_{GD} \\ \hat{a}_{GS+} \\ \hat{a}_{GD+} \\ \hat{a}_{YS+} \end{bmatrix} \sim N \left( \begin{bmatrix} \mu_{GS} \\ \mu_{GD} \\ \mu_{YS} \\ \mu_{GS} \\ \mu_{GD} \\ \mu_{GS+} \\ \mu_{GD+} \\ \mu_{YS+} \end{bmatrix}, \begin{bmatrix} 1 & 0 & \frac{1}{2} & R_{GS} & 0 & R_{GS+} & 0 & \frac{1}{2}R_{GS+} \\ & 1 & \frac{1}{2} & 0 & R_{GD} & 0 & R_{GD+} & \frac{1}{2}R_{GD+} \\ & & 1 & \frac{1}{2}R_{GS} & \frac{1}{2}R_{GD} & \frac{1}{2}R_{GS+} & \frac{1}{2}R_{GD+} & R_{YS+} \\ & & & R_{GS} & 0 & R_{GS} & 0 & \frac{1}{2}R_{GS+} \\ & & & & R_{GD} & 0 & R_{GD} & \frac{1}{2}R_{GD+} \\ & & & & & R_{GS+} & 0 & \frac{1}{2}R_{GS+} \\ & & & & & & R_{GD+} & \frac{1}{2}R_{GD+} \\ & & & & & & & R_{YS+} \end{bmatrix} \sigma_a^2 \right).$$

Using Pearson's results (Pearson, 1903) for multivariate normal distributions, GEBV and TBV were jointly simulated conditionally on the GS and GD EBV. These 2 EBV and their associated reliabilities were computed from the second BLUP evaluation on the real data set (i.e., in absence of progeny information of their sons). It follows that:

$$\begin{bmatrix} a_{GS} \\ a_{GD} \\ a_{YS} \\ \hat{a}_{GS+} \\ \hat{a}_{GD+} \\ \hat{a}_{YS+} \end{bmatrix} \sim N \left( \begin{bmatrix} \hat{a}_{GS} \\ \hat{a}_{GD} \\ \frac{\hat{a}_{GS} + \hat{a}_{GD}}{2} \\ \hat{a}_{GS} \\ \hat{a}_{GD} \\ \frac{\hat{a}_{GS} + \hat{a}_{GD}}{2} \end{bmatrix}, \mathbf{V} = \begin{bmatrix} \delta_{GS} & 0 & \frac{1}{2}\delta_{GS} & \Delta_{GS} & 0 & \frac{1}{2}\Delta_{GS} \\ 0 & \delta_{GD} & \frac{1}{2}\delta_{GD} & 0 & \Delta_{GD} & \frac{1}{2}\Delta_{GD} \\ \frac{1}{2}\delta_{GS} & \frac{1}{2}\delta_{GD} & \delta_{PED} & \frac{1}{2}\Delta_{GS} & \frac{1}{2}\Delta_{GD} & R_{YS+} - R_{PED} \\ \Delta_{GS} & 0 & \frac{1}{2}\Delta_{GS} & \Delta_{GS} & 0 & \frac{1}{2}\Delta_{GS} \\ 0 & \Delta_{GD} & \frac{1}{2}\Delta_{GD} & 0 & \Delta_{GD} & \frac{1}{2}\Delta_{GD} \\ \frac{1}{2}\Delta_{GS} & \frac{1}{2}\Delta_{GD} & R_{YS+} - R_{PED} & \frac{1}{2}\Delta_{GS} & \frac{1}{2}\Delta_{GD} & R_{YS+} - R_{PED} \end{bmatrix} \sigma_a^2 \right),$$

with  $\delta = 1 - R$  and  $\Delta = R_+ - R$ ;  $R$  represents the reliability based on records and relationships information only, whereas  $R_+$  refers to the genomically enhanced reliability. In practice, using a Cholesky decomposition of the variance-covariance matrix  $\mathbf{V}$ , simulated GEBV and TBV were obtained as linear combinations of 6 random standard normal variables. Consistent simulation of GEBV and TBV of full siblings was guaranteed through an

adequate repetition of the relevant random variables. Finally, the control and GPS populations differed in the way the breeding values of young sires were distributed. In the control population, random variables were drawn only once in such a way that the young sires' GEBV were normally distributed. No preselection existed. In contrast, in the GPS population, a genomic preselection step was implemented using a given selection rate defined as the proportion of retained candidates, say  $1/m$ . In practice, TBV and GEBV were generated, on average,  $m$  times for each YS, hence mimicking genomic evaluation of  $m$  full siblings. In order to be as realistic as possible, the number of full siblings within a family  $i$  ( $m_i$ ) was varied from one family to another. To ensure an expected number of  $m$  candidates with a minimum number of 1, a random number  $u_i$  was generated from a Poisson distribution  $P(\lambda)$  with  $\lambda = m - 1$ , and the size of the full-sibling cohort was set to  $u_i + 1$ . Among these full siblings, the highest GEBV was assigned to the YS undergoing progeny test, whereas the remaining full siblings were culled.

Performances for the daughters of each YS were simulated in the 2 populations. From the animal model equations, these records were computed as the sum of fixed and random effects. Without loss of generality, the only fixed effect ( $\beta$ ) considered was a contemporary group effect. In other words, data of the initial evaluation were precorrected for all nongenetic (fixed) effects and a contemporary group effect was estimated for each group of identified daughters. The breeding value of any daughter was generated as the sum of its average parental breeding value and a Mendelian sampling term ( $\varphi$ ) drawn from a normal distribution with zero mean and a variance equal to half the genetic variance of the trait. The TBV for the sire was available as described above, whereas the TBV for the dam was generated on the basis of her EBV and reliability computed from the second BLUP evaluation when no YS progeny performances were available. A normal random variable  $\varepsilon$  was, thus, added to dam EBV to get TBV. Finally, a residual  $e$  was drawn from a normal distribution with zero mean and a variance equal to the residual variance of the trait. Let  $y$  be the record of a daughter:

$$y = \beta + \frac{a_{YS} + (\hat{a}_D + \varepsilon)}{2} + \varphi + e$$

with

$$\left\{ \begin{array}{l} \varepsilon \sim N[0, (1 - R_D)\sigma_a^2] \\ \varphi \sim N(0, \frac{1}{2}\sigma_a^2) \\ e \sim N(0, \sigma_e^2) \end{array} \right\}.$$

## Numerical Applications

The simulations were first done following the national data set for a conformation trait, udder depth (**UD**), for the Holstein breed in France. A total of 4,110,229 records were available, and the pedigree file included 5,917,739 animals. The young sires were chosen among those born in 2001, 2002, and 2003, having more than 10 and fewer than 150 recorded daughters. In all, 1,875 sires fulfilled these criteria. Among them, 799 young sires were selected with their 40,222 daughters. Selection bias was first assessed for the UD trait with a heritability of 0.36, assuming that the top 25% (proportion,  $p = 0.25$ ) of the young sires were selected after genomic evaluation. The GEBV simulations also required choosing the number of daughters that would provide the same increase in reliability as that of the genomic evaluation. In the North American Holstein population, Van Raden et al. (2009) reported 9 daughter equivalents from genomic prediction for body depth with a heritability of 0.37. Here, an initial value of 10 genomic equivalent daughter contributions (**gEDC**) was chosen, which is equivalent to a reliability of genomic evaluation of 50%. This set of parameters ( $gEDC = 10$ ,  $p = 0.25$ ,  $h^2 = 0.36$ ) defines the udder depth trait reference scenario, **UD\_REF**.

## Sensitivity Analysis

In order to assess the magnitude of the bias due to genomic preselection and to understand the role of parameters such as the heritability of the trait, intensity of genomic selection, and reliability of genomic evaluations, various scenarios were implemented. Compared with that of the **UD\_REF** scenario, the value of heritability was decreased ( $h^2 = 0.14$ ): simulations were performed for another conformation trait, foot angle (**FA**). Notice that decreasing heritability makes the genomic evaluation less accurate if gEDC is kept equal to 10. The gEDC was increased to 26, which is possibly optimistic but made to maintain the same level of genomic accuracy. Even then, the cohorts of YS were no longer the same for FA. To keep fairly comparable scenarios, whereas handled traits are different, the YS for the FA trait were chosen so that the same average pedigree index was observed in both situations. A total of 601 young sires were, thus, identified with their 31,976 daughters. It defined the scenario hereafter called the foot angle trait reference scenario, **FA\_REF**. Then the proportion of selected animals was decreased from 0.25 to 0.10 for both traits, defining scenarios udder depth trait and foot angle trait, each with proportion of selected animals at 0.10 (**UD\_p** and **FA\_p**, respectively). Finally, for the FA trait, 2



**Table 2.** Parameters for the different studied scenarios

Scenario <sup>1</sup>	Proportion of selected candidates	Heritability	Genomic equivalent daughter contribution	Young sires (n)
UD_REF	0.25	0.36	10	799
UD_p	0.10	0.36	10	799
FA_REF	0.25	0.14	26	601
FA_p	0.10	0.14	26	601
FA_gEDC	0.25	0.14	10	601

<sup>1</sup>UD = udder depth trait; FA = foot angle trait; REF = reference; p = proportion of young sires retained; gEDC = genomic equivalent daughter contribution.

levels of genomic accuracy were compared. In the scenario called **FA\_gEDC**, gEDC was decreased from 26 (in FA\_REF) to 10. Reliability of direct genomic values decreased from 50% to 27%. In order to sum up (see Table 2), UD\_REF and FA\_REF were compared to assess the effect of heritability change. Having UD\_REF and UD\_p on one side, and FA\_REF and FA\_p on the other side, highlighted the effect of selection intensity. The comparison of FA\_REF with FA\_gEDC showed the importance of genomic accuracy.

**Criteria for Assessing the Quality of the Genetic Evaluations**

Bias was measured as the difference between TBV and EBV. Many criteria exist to assess the precision of evaluations; we retained 2 of them. The mean square error of prediction was preferred to the prediction error variance, as used in other studies on selection processes (Sorensen and Kennedy, 1984; Van Vleck, 1987; Schenkel et al., 2002). Hickey et al. (2008) and L. R. Schaeffer (Department of Animal and Poultry Science, University of Guelph, Canada, personal communication) considered the computation of the correlation between true and EBV. We decided to use the squared correlation to compare it to the approximate reliability, a byproduct of the mixed model equations. This reliability is computed using the standard deviation in the base population. Statistics were averaged over 50 replicates. Definitions of all criteria assessing

the quality of genetic evaluations are summed up in Table 3.

**RESULTS**

The results are first presented for the reference scenario (UD\_REF). To facilitate the interpretation, breeding values are expressed in genetic standard deviation of the trait.

**Validation of the Approach**

When the selection scheme involves progeny testing, information is collected for the whole population of selection candidates so that the estimation of the Mendelian sampling term is consistent with the hypotheses underlying the BLUP methodology. In particular, the expected value of the Mendelian sampling term is supposed to be zero. For both populations simulated in the UD\_REF scenario, Table 4 displays genetic values (TBV, EBV, MS) and the difference between EBV and TBV. In the control population, EBV are not significantly different from TBV and the MS term is not significantly different from 0. Hence, the considered population structure and estimation method lead to unbiased evaluations in classical conditions. And, in the GPS population, TBV are indeed significantly larger than the TBV observed in the control population (Table 4), illustrating the fact that genomic preselection was effective.

**Table 3.** Definitions of the criteria for assessing quality of genetic evaluations with replicate *r* and  $a_r = TBV$ ,  $\hat{a}_r = EBV$ <sup>1</sup>

Bias ( $\Delta$ )	Mean square error (MSE)	Squared correlation ( $\rho^2$ )
$= \frac{1}{50} \sum_{r=1}^{50} E(\hat{a}_r - a_r)$	$= \frac{1}{50} \sum_{r=1}^{50} [Var(\hat{a}_r - a_r) + E(\hat{a}_r - a_r)^2]$	$= \left[ \frac{1}{50} \sum_{r=1}^{50} \frac{cov(\hat{a}_r, a_r)}{\sqrt{var(\hat{a}_r) var(a_r)}} \right]^2$

<sup>1</sup>TBV = true breeding value; E = expected value; Var = variance.

**Table 4.** Mean (first line) and standard deviation within replicate (second row, in italics) of genetic values and bias ( $\Delta$ ) in the cohort of selection candidates when genomic selection is implemented (GPS) or not (control)<sup>1</sup>

Population	TBV	EBV	MS	$\Delta = \overline{EBV - TBV}$
Control	0.931 <i>0.864</i>	0.933 <i>0.742</i>	-0.001 (NS) <i>0.630</i>	0.002 (NS) <i>0.427</i>
GPS	1.384 <i>0.781</i>	1.238 <i>0.694</i>	0.304*** <i>0.578</i>	-0.146 *** <i>0.409</i>

<sup>1</sup>TBV = true breeding value; MS = Mendelian sampling contribution. Values are averaged over 50 replicates. Tested the null hypothesis  $H_0$ , where the mean of MS or mean of bias equal zero.

\*\*\* $P < 0.001$ .

### Bias Evidence

Table 4 indicates that the MS term in the GPS population is significantly larger than zero, whereas the variance has decreased compared with that in the control population. From this first scenario, the basic hypothesis on the MS distribution is shown to be no longer consistent with the observed distribution. Finally, the difference between EBV and TBV is significant and negative ( $\Delta = -0.146$ ), indicating that evaluations are biased when genomic preselection is implemented. Furthermore, the expected value of the bias ( $\Delta$ ) between the 50 replicates presents a low variability in the control (standard deviation,  $\sigma_\Delta = 0.016$ ) and in the GPS ( $\sigma_\Delta = 0.014$ ) populations, showing that the results on bias are not random.

### Distribution of the Bias in Genetic Evaluations

Under the classical selection scheme, the difference between EBV and TBV is distributed around a mean value of 0 with a standard deviation within replicate of 0.427, as indicated in Tables 4 and 5. This is a symmetric distribution where mean and median are superposed, the distance between third and second, and second and first quartiles are the same (Table 5). It displays the distribution of the prediction error which is minimized with BLUP. When genomic preselection is implemented, such a type of distribution is also observed, but values

are systematically translated into more negative values. On average, bias is equal to  $-0.146$  with a standard deviation between replicates of 0.014 in the YS cohort and bias is equal to  $-0.044$  with a standard deviation between replicates of 0.006 in the cohort of their daughters. Hence, EBV underestimate TBV.

### Effect on the Accuracy of Evaluations

By construction, the BLUP solutions should maximize the correlation between TBV and EBV and minimize the mean square error of prediction. The MSE increased in the GPS population compared with those of the control population. Similarly, the squared correlation between TBV and EBV was lower in the GPS population than that in the control population (Table 6), whereas the amount of information was identical, as indicated by the value of the approximate reliability (REL). All of the indicators show that the genomic preselection step decreased the accuracy of the estimations compared with that of the control population, with a diluted effect on daughters compared with that of young sires.

### Sensitivity Analysis

Table 7 indicates that all parameters (intensity of selection, heritability, and accuracy of genomic evaluation) induced changes in the mean bias and mean MSE

**Table 5.** Quartiles and first moments of the difference between EBV and true breeding values for the young sires (YS) and their daughters (D) when genomic selection is implemented (GPS) or not (control)<sup>1</sup>

Item	Control population		GPS population	
	YS	D	YS	D
Minimum	-2.249	-3.523	-2.598	-3.655
First quartile	-0.272	-0.486	-0.406	-0.539
Median	0.004	0.011	-0.138	-0.043
Mean	0.002 (NS)	0.011**	-0.146***	-0.044***
SD between replicates	0.016	0.006	0.014	0.006
Third quartile	0.278	0.508	0.127	0.454
Maximum	2.195	3.700	1.915	3.608
SD within replicate	0.427	0.737	0.409	0.736

<sup>1</sup> Tested the null hypothesis  $H_0$ , where the mean of MS or mean of bias equal zero. \*\* $P < 0.01$ ; \*\*\* $P < 0.001$ .



**Table 6.** Mean squared error (MSE) and reliability measures [ $\rho^2$ (EBV, TBV), REL] in control and GPS populations for udder depth trait and 25% selection rate<sup>1</sup>

Population	MSE	$\rho^2$ (EBV, TBV)	REL
Control			
Young sires	0.183	0.756	0.815
Daughters	0.544	0.414	0.476
GPS			
Young sires	0.188	0.727	0.815
Daughters	0.544	0.394	0.476

<sup>1</sup> $\rho^2$  = squared correlation; TBV = true breeding value; REL = approximate reliability; GPS = genomic pre-selected sires.

of the BLUP evaluations compared with those of the reference scenario, UD\_REF. When genomic selection intensity increased, the magnitude of the bias also rose and the accuracy of the estimations decreased. This is verified for UD and FA traits when the proportion of selected candidates decreased from 25 to 10% (scenario UD\_REF compared with UD\_p and scenario FA\_REF compared with FA\_p). The same trends were observed when the precision of the genomic evaluation decreased (FA\_gEDC compared with FA\_REF). Finally, scenarios UD\_REF and FA\_REF were compared, with the same proportion of selected candidates and the same amount of genomic information. When the heritability was lower, genetic evaluations were more affected: the magnitude of the bias was larger and the accuracy of the estimations was lower.

## DISCUSSION

Simulating data such as TBV and phenotypic values has many advantages. Compared with studies on real data, the only relevant source of bias is isolated and the mechanisms underlying bias are controlled so that it is then easy to vary parameters of interest and better understand mechanisms underlying bias. In the literature, several methods were proposed to systematically detect bias in prediction procedures. They involved assessing the change between estimates of 2 sets of selected and unselected data to study sequential selection (Mallinckrodt et al., 1995) or between 2 consecutive predictions (Reverter et al., 1994). However, by accessing true genetic values, the actual properties of BLUP solutions may be contrasted with their expectation of unbiasedness and of minimum error variance. The major limitation with this method is that it cannot be used to check routinely if classical evaluations are biased by genomic preselection.

The approach in our study was designed to be realistic. It is based on real data with large numbers of herds and young sires promoting genetic and environmental diversity. Moreover, the choice of paternal families and their composition (number of siblings) tried to mimic

the strategies implemented by breeding companies when genotyping candidates. However, only one data structure was considered with type traits. Repeated or longitudinal data, for example, for which evaluation models are different, were not investigated.

Genomic selection can be seen as a missing data process where only progeny of selected animals are recorded. These were selected based on GEBV; however, this information was not included in the evaluation. Actually, whatever the scenario, it is always observed that under genomic selection, bias is significant and breeding values of selected young sires and their daughters are underestimated. Mixed model equations assume that selected young sires have a Mendelian sampling term with zero expectation, whereas the MS contribution is now, on average, positive. The accuracy of evaluations is also always lower when a genomic preselection step is implemented. The magnitude of the bias and the accuracy loss are even greater when the intensity of genomic selection is increasing. This is not surprising, because the proportion of missing data is indeed rising. We observe the same trends when the accuracy of the genomic evaluations is decreasing, when the correlation between GEBV and the trait is in the same way declin-

**Table 7.** Mean bias ( $\Delta$ ), standard deviation ( $\sigma_\Delta$ ), and mean squared error (MSE) over 50 replicates when the heritability, the proportion of selected candidates, and the reliability of genomic evaluation vary

Scenario <sup>1</sup>	$\Delta, \sigma_\Delta$	MSE
UD_REF	-0.146 $\pm$ 0.014	0.188
UD_p	-0.227 $\pm$ 0.016	0.217
FA_REF	-0.214 $\pm$ 0.021	0.305
FA_p	-0.155 $\pm$ 0.020	0.299
FA_gEDC	-0.249 $\pm$ 0.025	0.364

<sup>1</sup>UD = udder depth trait; FA = foot angle trait; REF = reference; p = proportion of young sires retained; gEDC = genomic equivalent daughter contribution; UD\_REF = 36% of heritability, 25% of young sires retained, 10 EDC from genomic evaluation; UD\_p = 36% of heritability, 10% of young sires retained, 10 EDC from genomic evaluation; FA\_REF = 14% of heritability, 25% of young sires retained, 26 EDC from genomic evaluation; FA\_p = 14% of heritability, 10% of young sires retained, 26 EDC from genomic evaluation; FA\_gEDC = 14% of heritability, 10% of young sires retained, 10 EDC from genomic evaluation.

ing. Finally, Sorensen and Kennedy (1984) reported that the magnitude of a selection bias also depended on heritability. The coefficients of the relationship matrix being incorrect and even combined with the variance ratio in the mixed model equations may explain how bias and MSE vary with heritability.

The magnitude of the bias was significantly different from zero under genomic selection in all of the studied scenarios. Over one generation only, the bias comprises between 4 and 11% of genetic standard deviation for the daughters of the young sires, and between 15 and 25% for the young sires. For the same range of heritability and selection intensity, such magnitude was never reported in studies assessing bias in genetic evaluations. As genomic evaluations are predicted to be the future dominant selection tool (progeny testing is already being abandoned in some breeding programs), the preselection intensity, which is not yet accounted for in classical models, may become large, and clearly biased evaluations are to be feared. Henderson (1990a, b) reported that a biased predictor may exist that has a smaller mean squared error than an unbiased predictor. Indeed, Gianola et al. (1988) and L. R. Schaeffer (Department of Animal and Poultry Science, University of Guelph, Canada, personal communication) used this argument to stress that the most important aspect to increase genetic gain is the increase of EBV accuracy. However, our results indicate an increase of the MSE. Assuming genomic preselection of the young sires only, genetic progress may then be decreased by 3 to 4%, if 25 or 10% of YS are retained. It means that an uncontrolled loss of accuracy can actually threaten the effectiveness of a breeding program and have large genetic and then economic effects. It seems essential to take into account this bias. To assess the bias magnitude over more generations, a fully stochastic study would have to be used. Considering previous studies, we expect that effect on genetic evaluations would increase. For example, Schenkel et al. (2002) found that the bias increased over generations in the case of missing pedigree information. In fact, direction and amplitude of the bias may particularly become unpredictable when daughters of selected young sires are mated to other preselected young sires.

## CONCLUSIONS

This study presents evidence that national genetic evaluations are biased once young sires are preselected based on genomic breeding values. The challenge is now to account for genomic preselection in the classical evaluation models. Understanding why evaluations are biased leads us to propose the inclusion of, in the classical evaluation, performance obtained from

deregressed GEBV of culled and selected candidates, whereas Misztal et al. (2009) proposed a single-step genetic evaluation combining pedigree-based information with genomic data. All information upon which selection has been based would then be included. In addition, genomic and polygenic information would be naturally combined so that the level of accuracy of national evaluations would be increased. However, a correction, at a national level, would be even more relevant for international evaluations where breeding practices strongly differ between countries. At least, a validation test would be required to check that national evaluations are still unbiased despite genomic preselection to ensure the quality of international evaluations.

## ACKNOWLEDGMENTS

Financing from Agence Nationale de la Recherche project AMASGEN (Jouy-en-Josas, France) and APIS-GENE are acknowledged. The authors thank members of the AMASGEN project for their comments: P. Croiseau, S. Fritz, F. Guillaume, A. Legarra, C. Robert-Granie.

## REFERENCES

- Ducrocq, V., M. Goddard, L. Schaeffer, and G. Banos. 2008. Interbull Scientific Advisory Committee Annual Report (2007–2008) to the Interbull Steering Committee. 4 pp. Interbull meeting, Niagara Falls, NY.
- Gianola, D., S. Im, and R. L. Fernando. 1988. Prediction of breeding value under Henderson's selection model: A revisitation. *J. Dairy Sci.* 71:2790–2798.
- Harris, B., and D. Johnson. 1998. Approximate reliability of genetic evaluations under an animal model. *J. Dairy Sci.* 81:2723–2728.
- Henderson, C. R. 1990a. Accounting for selection and mating biases in genetic evaluation. Pages 413–436 in *Advances in Statistical Methods for Genetic Improvement of Livestock*. D. Gianola and K. Hammond, ed. Springer-Verlag, Heidelberg, Germany.
- Henderson, C. R. 1990b. Statistical methods in animal improvement: Historical overview. Pages 2–14 in *Advances in Statistical Methods for Genetic Improvement of Livestock*. D. Gianola and K. Hammond, ed. Springer-Verlag, Heidelberg, Germany.
- Hickey, J. M., M.G. Keane, D. A. Kenny, A. R. Cromie, H. A. Mulder, and R. F. Veerkamp. 2008. Estimation of accuracy and bias in genetic evaluations with genetic groups using sampling. *J. Anim. Sci.* 86:1047–1056.
- Loberg, A., and J. Dürr. 2009. Interbull survey on the use of genomic information. Pages 3–14 in *Proc. Interbull International Workshop on Genomic Information in Genetic Evaluations*. Interbull Bulletin No. 39. Interbull, Uppsala, Sweden.
- Mallinckrodt, C. H., B. L. Golden, and R.M. Bourdon. 1995. The effect of selective reporting on estimates of weaning weight parameters in beef cattle. *J. Anim. Sci.* 73:1264–1270.
- Misztal, I., A. Legarra, and I. Aguilar. 2009. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. *J. Dairy Sci.* 92:4648–4655.
- Pearson, K. 1903. Mathematical contributions to the theory of evolution. XI. On the influence of natural selection on the variability and correlation of organs. *Philos. Trans. Royal Soc. London* 200:1–66.
- Reverter, A., B. L. Golden, R. M. Bourdon, and J. S. Brinks. 1994. Technical note: Detection of bias in genetic predictions. *J. Anim. Sci.* 72:34–37.

- Schenkel, F. S., L. R. Schaeffer, and P. J. Boettcher. 2002. Comparison between estimation of breeding values and fixed effects using Bayesian and empirical BLUP estimation under selection on parents and missing pedigree information. *Genet. Sel. Evol.* 34:41–59.
- Sorensen, D. A., and B. W. Kennedy. 1983. The use of the relationship matrix to account for genetic drift variance in the analysis of genetic experiments. *Theor. Appl. Genet.* 66:217–220.
- Sorensen, D. A., and B. W. Kennedy. 1984. Estimation of response to selection using least-squares and mixed model methodology. *J. Anim. Sci.* 58:1097–1106.
- van der Beek, S. 2007. Effect of genomic selection on national and international genetic evaluations. Pages 115–118 in *Proc. Interbull Meeting: Dublin, Ireland*. Interbull Bulletin No. 37. Interbull, Uppsala, Sweden.
- VanRaden, P. M., C. P. Van Tassell, G. R. Wiggans, T. S. Sonstegard, R. D. Schnabel, J. F. Taylor, and F. S. Schenkel. 2009. *Invited review: Reliability of genomic predictions for North American Holstein bulls*. *J. Dairy Sci.* 92:16–24.
- Van Vleck, L. D. 1987. Contemporary groups for genetic evaluations. *J. Dairy Sci.* 70:2456–2464.