Journal of Dairy Research

cambridge.org/dar

Comparison of different response variables in genomic prediction using GBLUP and ssGBLUP methods in Iranian Holstein cattle

Mohamadreza Afrazandeh[1], Rostam Abdolahi-Arpanahi[2], Mokhtar Ali Abbasi[3], Nasser Emam Jomeh Kashan[1] and Rasoul Vaez Torshizi[4]

[1]Department of Animal Science, Faculty of Agriculture Sciences and Food Industries, Science and Research Branch, Islamic Azad University, Tehran, Iran; [2]Department of Animal and Dairy Science, College of Agricultural and Environmental Sciences, University of Georgia, Athens, USA; [3]Animal Science Research Institute of Iran, Agricultural Research, Education and Extension Organization (AREEO), Karaj, Iran and [4]Department of Animal Science, Faculty of Agriculture, Tarbiat Modares University, Tehran, Iran

Research Article

Cite this article: Afrazandeh M, Abdolahi-Arpanahi R, Abbasi MA, Kashan NEJ and Torshizi RV (2022). Comparison of different response variables in genomic prediction using GBLUP and ssGBLUP methods in Iranian Holstein cattle. *Journal of Dairy Research* **89**, 121–127. https://doi.org/10.1017/S0022029922000395

Received: 28 July 2021
Revised: 16 February 2022
Accepted: 28 February 2022
First published online: 23 May 2022

Keywords:
Daughter yield deviation; de-regressed proof; genomic selection; response variable

Author for correspondence:
Rostam Abdolahi-Arpanahi,
Email: rostam.abdollahi@uga.edu

Abstract

We compared the reliability and bias of genomic evaluation of Holstein bulls for milk, fat, and protein yield with two methods of genomic best linear unbiased prediction (GBLUP) and single-step GBLUP (ssGBLUP). Four response variables of estimated breeding value (EBV), daughter yield deviation (DYD), de-regressed proofs based on Garrick (DRP_{GR}) and VanRaden (DRP_{VR}) were used as dependent variables. The effects of three weighting methods for diagonal elements of the incidence matrix associated with residuals were also explored. The reliability and the absolute deviation from 1 of the regression coefficient of the response variable on genomic prediction (Dev) using GBLUP and ssGBLUP methods were estimated in the validation population. In the ssGBLUP method, the genomic prediction reliability and Dev from un-weighted DRP_{GR} method for milk yield were 0.44 and 0.002, respectively. In the GBLUP method, the corresponding measurements from un-weighted EBV for fat were 0.52 and 0.008, respectively. Moreover, the un-weighted DRP_{GR} performed well in ssGBLUP with fat yield values for reliability and Dev of 0.49 and 0.001, respectively, compared to equivalent protein yield values of 0.38 and 0.056, respectively. In general, the results from ssGBLUP of the un-weighted DRP_{GR} for milk and fat yield and weighted DRP_{GR} for protein yield outperformed other models. The average reliability of genomic predictions for three traits from ssGBLUP was 0.39 which was 0.98% higher than the average reliability from GBLUP. Likewise, the Dev of genomic predictions was lower in ssGBLUP than GBLUP. The average Dev of predictions for three traits from ssGBLUP and GBLUP were 0.110 and 0.144, respectively. In conclusion, genomic prediction using ssGBLUP outperformed GBLUP both in terms of reliability and bias.

Genomic selection has been adopted as a standard tool for genetic evaluation in different livestock species (Misztal *et al*., 2020). In genomic selection, there is a need to establish a reference population comprised of animals with phenotypic records and genotypes for single-nucleotide polymorphism (SNP) markers. Usually, the target phenotypic values of the reference population are called response variables. Historically, the estimated breeding values (EBVs) of animals were predicted from the phenotypic and pedigree information and then in the reference population, the EBVs were used as response variables to estimate the effects of SNPs. The genomic estimated breeding values (GEBV) of animals were predicted from the sum of these effects. Finally, the candidate animals were selected according to their GEBVs. Nowadays, the methodology of genomic selection benefits from the construction of a genomic relationship matrix (**G**) (VanRaden, 2008), which consists of genomic relationships of the animals and allows conceptual comparisons between pedigree-based and genome-based predictions. Meanwhile, the information regarding the relations of a dairy bull is used for the prediction of his EBV and also for creating the **G** matrix. Therefore, if the EBV is used as a response variable, it causes double-counting of information (Garrick *et al*., 2009).

To avoid double-counting of information derived from the bull's relatives, the daughter yield deviation (DYD) and de-regressed proofs (DRP) are proposed as response variables. The DYD of a bull is derived from the information of its daughters only and is the average performance of the daughters adjusted for all fixed effects as well as the EBV of mates of the bull (Mrode, 2014). Therefore, using DYD as a response variable for genomic evaluation does not cause the problem of double-counting of information. The DRP is estimated from the EBV, which is obtained by dividing the EBV of the animal by its reliability (Goddard, 1985) or setting up the complete mixed model equations (MME) for all animals in the pedigree (Jairath *et al*., 1998). Thus, to simplify and prevent reconstructing the complete MME, simpler

strategies have been proposed (Garrick *et al.*, 2009; VanRaden *et al.*, 2009). In genomic evaluation of animals, comparing these different response variables from the bias point of view is an important issue. In a study on Chinese Holstein cattle, the results of using the EBV, DRP_{GR} (Garrick *et al.*, 2009) and DRP_{JR} (Jairath *et al.*, 1998) as response variables for genomic evaluation were compared. The DRP_{JR} outperformed DRP_{GR} and EBV in terms of accuracy and unbiasedness (Song *et al.*, 2018).

Genomic breeding values can be predicted with one of two models of GBLUP or ssGBLUP. The GBLUP has the same form as the conventional BLUP model, but the inverse of the numerator relationship matrix ($\mathbf{A}^{-1}$) is replaced by the inverse of the genomic relationship matrix ($\mathbf{G}^{-1}$) (Hayes *et al.*, 2009). Using the genomic relationship, the proportion of chromosome segments shared by individuals can be estimated. The reason is that high-density genotyping identifies genes identical in state (Forni *et al.*, 2011). However, in the ssGBLUP approach, the **A** matrix is replaced by the **H** matrix which combines **G** and **A** matrices (Misztal *et al.*, 2009). In some cases, **A** and **G** matrices are not on the same scale (Misztal, 2017) and one needs to use optimal scaling factors to blend $\mathbf{G}^{-1}$ with the inverse of the pedigree relationship matrix ($\mathbf{A}_{22}^{-1}$) for the genotyped animals (Vitezica *et al.*, 2011; Misztal *et al.*, 2013).

Generally, the reliability of predicted response variables is not the same. Thus, in genomic predictions, the weighted analyses are carried out to account for heterogeneous residual variances among bulls due to differences in reliabilities of response variables. Therefore, the performance of ssGBLUP and GBLUP can be affected not only by the type of response variable but also by weighting of residuals. In a study of genomic prediction, in which the DRP was used as the response variable, the reliability of GEBV from ssGBLUP was 2.1% higher than reliability from GBLUP (Gao *et al.*, 2012). In another study, the use of ssGBLUP based on the DRP response variable led to slightly higher reliability than GBLUP (Koivula *et al.*, 2012). Thus, this study aimed to estimate the reliability of genomic prediction with two methods of GBLUP and ssGBLUP, using four response variables including EBV, DYD, and two DRPs with weighted and un-weighted residuals. We used the data of Iranian Holstein dairy cattle.

Materials and methods

Data

The phenotypic performance, pedigree, and genotypes of Iranian Holstein cattle were provided by the Animal Breeding Center of Iran. The dataset consisted of 651 985 and 479 268 and 425 151 records of milk (MY), fat (FY) and protein (PY) yield, respectively, from cows of sires born in the years 1989–2014. There were 101 bulls genotyped with low-density (<20 k), 749 bulls with medium-density (>20 and <60 k), and 759 bulls with high-density (>60 k) SNP chips. SNPs with minor allele frequency (MAF) less than 0.01, call rate for each marker less than 0.95 and Hardy-Weinberg equilibrium less than α/n (α equals 0.05 and n is the number of SNPs) were removed by the QCf90 software (Misztal *et al.*, 2002). Then, all genotypes were imputed to 40 k by FImpute software (Sargolzaei *et al.*, 2014). Finally, 1609 genotyped bulls and 41 135 SNPs were retained for analysis.

The de-regressed proof was estimated by two methods of de-regression, namely those of Garrick *et al.* (2009) and VanRaden *et al.* (2009). The DYD was calculated as described in Mrode (2014). Then, they were used as response variables for GEBV prediction. The two de-regression methods and DYD which were used in this study calculated as follows:

Vanraden method (DRP_{VR}).

$$DRP_i = \left[\frac{EBV_i - PA_i}{R_i^2}\right] + PA_i$$

where DRP_i is DRP for bull i, PA_i average EBV of parents of bull i, EBV_i is the estimated breeding value of bull i, R_i^2 is reliability of DRP_i and is calculated with:

$$R_i^2 = \frac{ERC_{P_i}}{ERC_{P_i} + ERC_{PA_i} + 1}$$

where ERC_{P_i} is the effective record contributions of progeny for the bull i, ERC_{PA_i} is the effective record contributions for parent of bull i. They were estimated with the following formulae (VanRaden and Wiggans, 1991):

$$ERC_{P_i} = \left[\lambda \frac{REL_{EBV_i}}{(1 - REL_{EBV_i})}\right] - ERC_{PA_i}$$

$$ERC_{PA_i} = \lambda \frac{REL_{PA_i}}{(1 - REL_{PA_i})}$$

where $\lambda = (1 - h^2)/h^2$, REL_{EBV_i} is reliability of EBV_i and REL_{PA_i} is reliability of PA_i.

Garrick method (DPR_{GR})

The equations solved to get DRP for each animal are as follows (Garrick *et al.*, 2009):

$$\begin{bmatrix} Z_{PA}'Z_{PA} + 4\lambda & -2\lambda \\ -2\lambda & Z_i'Z_i + 2\lambda \end{bmatrix} \begin{bmatrix} PA \\ EBV \end{bmatrix} = \begin{bmatrix} y_{PA}^* \\ y_i^* \end{bmatrix}$$

The elements of the matrices are:

$$Z_{PA}'Z_{PA} = \lambda(0.5\alpha - 4) + 0.5\lambda\sqrt{(\alpha^2 + 16/\delta)}$$

$$\alpha = 1/(0.5 - R_{PA}^2)$$

$$Z_i'Z_i = \delta Z_{PA}'Z_{PA} + 2\lambda(2\delta - 1)$$

$$\delta = (0.5 - R_{PA}^2)/(1 - R_{EBV}^2)$$

$$y_i^* = -2\lambda\, PA_i + (Z_i'Z_i + 2\lambda)EBV_i$$

$$DRP_i = (y_i^*/Z_i'Z_i) + PA_i$$

The reliability of DRP was calculated as:

$$R_{DRP_i}^2 = 1 - \lambda/(Z_i'Z_i + \lambda)$$

https://doi.org/10.1017/S0022029922000395 Published online by Cambridge University Press

Daughter yield deviation method (DYD)

The DYD was calculated as below (Mrode, 2014).

$$DYD_i = \frac{\sum_i^k u_{prog} * n_{prog} * (2YD_{prog} - \hat{a}_{mate})}{\sum_i^k (u_{prog} * n_{prog})} + PA_i$$

where k is the number of daughters of the bull i, YD_{prog} is the deviation of performance of daughters of the bull i from the average of population. It adjusts production of daughters of the bull i for all effects (except for additive animal genetic and error effect) and $\hat{a}_{mate}$ is breeding value of the bull i mate. If the mate of the bull i is known the u_{prog} is 1 and if it is not known equals 2/3. The reliability of DYD was calculated with the following formulae (VanRaden *et al.*, 2009).

$$R^2_{DYD} = \frac{DE_{prg}}{DE_{prg} + 1}$$

$$DE_{prg} = \frac{R^2_{EBV}}{1 - R^2_{EBV}} - \frac{R^2_{PA}}{1 - R^2_{PA}}$$

where R^2_{PA} is the reliability of parent average EBV of the bull i ($R^2_{sire} + R^2_{dam}/4$). The DE_{prg} is the daughter equivalent from daughters information.

GBLUP and ssGBLUP methods

Using the EBV, DYD, and two DRPs as response variables, the GEBV of bulls was predicted with GBLUP and ssGBLUP methods. The statistical model is as:

$$\boldsymbol{y} = 1\mu + \mathbf{Z}\boldsymbol{g} + \boldsymbol{e}$$

where $\boldsymbol{y}$ is the vector of response variable, μ is the total mean, $\boldsymbol{1}$ is the vector with all elements of 1, $\mathbf{Z}$ is incidence matrix which connects $\boldsymbol{g}$ to $\boldsymbol{y}$, $\boldsymbol{g}$ is the vector of additive genetic effects of all genotyped bulls and $\mathbf{e}$ is the vector of residuals. The additive genetic effects have a normal distribution with $N(0, \mathbf{G}\,\sigma^2_g)$, or $N(0, \mathbf{H}\sigma^2_g)$, σ^2_g is the additive genetic variance and $\mathbf{G}$ represents the genomic relationship matrix (VanRaden, 2008), and $\mathbf{H}$ is matrix which combines $\mathbf{G}$ and $\mathbf{A}$ matrices. The dimensions of matrices $\mathbf{G}$ and $\mathbf{H}$ were 1609 and 5133, respectively. $\boldsymbol{e}$ is the vector of random residuals with a normal distribution $N(0, \mathbf{D}\sigma^2_e)$, σ^2_e is residual variance and $\mathbf{D}$ represents a diagonal matrix with $b_{ii} = 1/W_i$, where W_i is the weight.

Since the $\mathbf{G}$ matrix was not positive definite, therefore 5% of $\mathbf{A}$ matrix was added to 95% of $\mathbf{G}$ matrix. The $\mathbf{H}$ matrix blends the pedigree and genomic information (Legarra *et al.*, 2009). The $\mathbf{H}^{-1}$ matrix is constructed as follows (Aguilar *et al.*, 2010; Christensen and Lund, 2010):

$$\mathbf{H}^{-1} = \begin{bmatrix} 0 & 0 \\ 0 & \tau(0.95\mathbf{G} - 0.05\mathbf{A}_{22})^{-1} - \omega\mathbf{A}^{-1}_{22} \end{bmatrix} + \mathbf{A}^{-1}$$

The $\mathbf{A}^{-1}$ is inverse of the pedigree-based relationship matrix and $\mathbf{A}^{-1}_{22}$ is the inverse of the subset of $\mathbf{A}$ for genotyped individuals. The $\mathbf{A}^{-1}$ consisted of individuals with genotype (1609 bulls) and the ancestors up to three generations ago. Therefore, the dimensions of matrix $\mathbf{A}$ were 5133×5133. The τ and ω as scaling factors were used for accounting for the reduced genetic variance and different depths of pedigree, respectively, to make $\mathbf{G}^{-1}$ compatible with $\mathbf{A}^{-1}_{22}$ and also $\mathbf{A}^{-1}$. Different values were tested for τ and ω, so that the optimal scaling factors had the lowest bias and the highest reliability for each response variable were selected.

Table 1. Number of bulls, progeny and the size of reference and validation populations for milk (MY), fat (FY) and protein (PY) yields in GBLUP and ssGBLUP methods

	Number of bulls		
Number of progeny	MY	FY	PY
10 ≤progeny <50	232	283	290
50 ≤progeny <150	275	273	273
150 ≤progeny <250	130	83	87
250 ≤progeny <350	68	81	79
progeny≥350	224	148	139
Reference population size	818	780	780
Validation population size	111	88	88

Weights

For GEBV predictions with two methods of GBLUP and ssGBLUP, the residual variance matrix was weighted with three different formulae:

(1) When EBV and DYD used as response variables, the diagonal elements of $\mathbf{D}$ matrix were weighted with $R^2/1 - R^2$, where the R^2 is the reliability of response variables. This weight is called as $W_{classic}$ in the context.
(2) When DRP_{GR} and DRP_{VR} were used as response variables, the ERCp was used as the weight (VanRaden and Wiggans, 1991).
(3) For DRP_{GR} and DRP_{VR}, a new formula was used for the estimation of the weight (Garrick *et al.*, 2009).

$$W_{GR_i} = (1 - h^2)/[(c + (1 - r^2_i)/(r^2_i)) \times h^2]$$

where, c is the proportion of genetic variance which is not captured by markers and c was 0.1 and r^2_i was calculated according to Garrick *et al.* (2009).

Genomic prediction

In this study, different datasets were prepared to assess prediction performance for three traits: (1) a full dataset containing all records of cows of which their sires born during years 1989–2014, and EBV, DYD and two DRPs were calculated for bulls to be used as benchmark; and (2) a reduced dataset included records of cows of which their sires born during years 1989–2012 and EBV, DYD and two DRPs were calculated for bulls to be used in genomic prediction. Subsequently, bulls from the reduced dataset were assigned to the reference population according to year of birth (1989–2012) and bulls born during 2013–2014 were assigned to validation population. The validation population (used to assess genomic prediction reliability and bias) included only genotyped bulls with no daughters in the reduced dataset, but with at least 10 daughters in the full dataset. Table 1 provides summary information such as number of bulls, the progeny, and the size of reference and validation populations for three traits in GBLUP and ssGBLUP methods.

https://doi.org/10.1017/S0022029922000395 Published online by Cambridge University Press

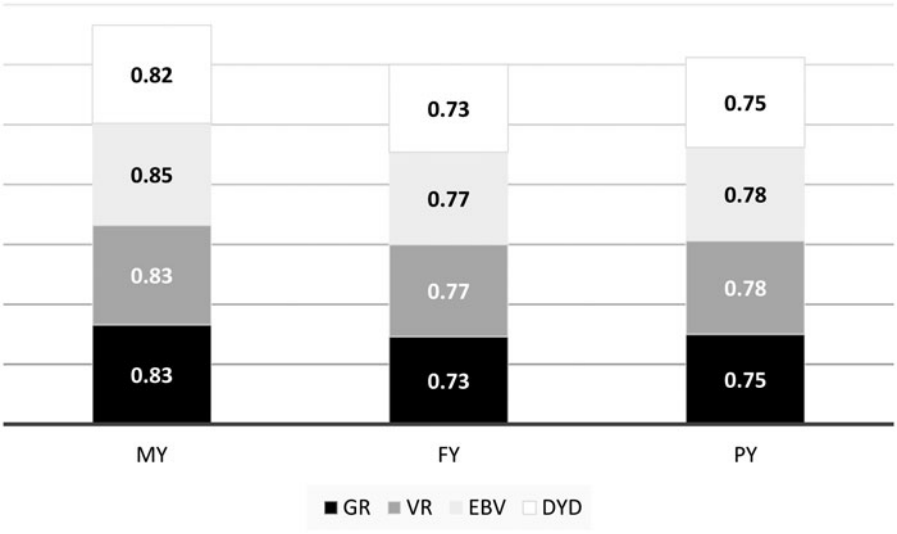

Fig. 1. The reliability of response variables for milk (MY), fat (FY) and protein (PY) yields. EBV, estimated breeding value; DYD, daughter yield deviation; VR, de-regressed proof estimated by VanRaden's formula; GR, de-regressed proof estimated by Garrick's formula.

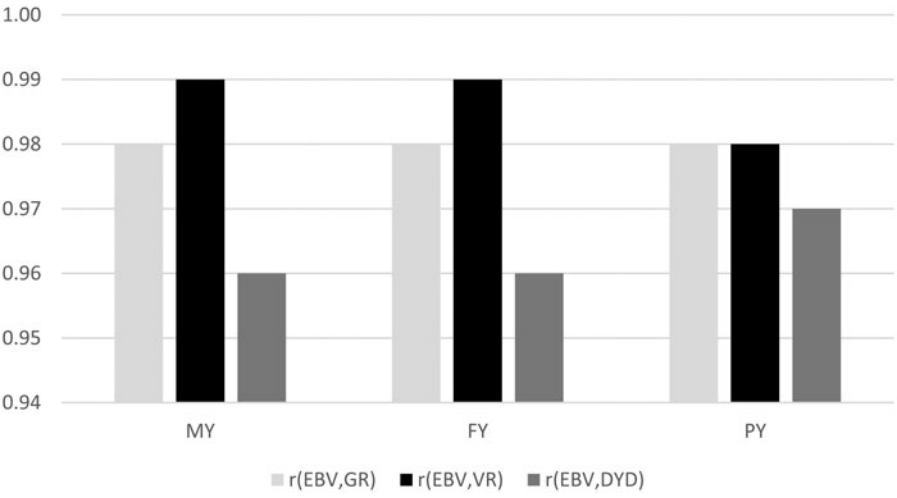

Fig. 2. The correlation of DYD, DRP_{GR} and DRP_{VR} with EBV for milk (MY), fat (FY) and protein (PY) yields. EBV, estimated breeding value; DYD, daughter yield deviation; VR, de-regressed proof estimated by VanRaden's formula; GR, de-regressed proof estimated by Garrick's formula.

Validation

The reliability of genomic predictions for the studied traits was measured as the squared correlation between genomic prediction (obtained with the reduced dataset) and response variable (EBV, DYD or two DRPs from the full dataset) divided by the average reliability of response variable in the validation datasets (Gao *et al.*, 2012).

To access the bias of genomic predictions for each method, the following regression model was used:

$$RV_i = b_0 + b_1 X_p i + e_i$$

where RV is the response variable (EBV, DYD or two DRPs), obtained from the full dataset, of the ith validation bull; b_0 is the intercept; $b_1$is the linear regression coefficient indicating bias (bias in dispersion) of the predictions; X_p is the ith bull's genomic prediction obtained from the reduced dataset; and e is the residual.

Results and discussion

Using the full dataset and the single-trait model, the estimates of heritability (±standard error) for MY, FY and PY were 0.30 (±0.003), 0.21 (±0.004) and 0.24 (±0.004), respectively. The estimates correspond with the results of another report which were 0.39, 0.29 and 0.31 for MY, FY and PY, respectively from Holstein cattle of Canada (Oliveira *et al.*, 2018). The average reliabilities of EBV and DRP_{VR} for three traits were 0.80 and 0.79, respectively. The response variable with the highest reliability was EBV followed by DRP_{VR}, DRP_{GR} and DYD (Fig. 1).

The highest estimated correlation ($r \sim 0.99$) was observed between DRP_{VR} and EBV and the lowest correlation ($r \sim 0.96$) was between DYD and EBV (Fig. 2). The average correlation between DRP_{VR} and EBV was 0.99 for three traits which showed the estimates of DRP_{VR} were almost the same as the estimates of the EBV (Fig. 2).

Comparison of different response variables in genomic prediction

The genomic prediction is affected by the accuracy of marker effects estimation which depends on the information in response variables. The reliability, bias and the absolute deviation from 1 of the regression coefficient of the response variable on genomic prediction (Dev) using GBLUP and ssGBLUP methods for the validation population are presented in Table 2. In ssGBLUP, the

https://doi.org/10.1017/S0022029922000395 Published online by Cambridge University Press

Table 2. Reliabilities (R^2), regression coefficients (b_1) and the absolute deviation of the regression coefficients from 1.0 (Dev) for three traits of milk (MY), fat (FY) and protein (PY) yields in GBLUP and ssGBLUP methods

			GBLUP			ssGBLUP		
Trait	Response variable	Weight	R^2	b_1	Dev	R^2	b_1	dev
Milk	EBV	$W_{without}$	0.3154	0.814	0.186	0.3418	0.812	0.188
	$W_{classic}$	0.3233	0.742	0.258	0.3419	0.812	0.188	
	DYD	$W_{without}$	0.3994	0.892	0.108	0.3995	0.892	0.108
		$W_{classic}$	0.4105	0.937	0.063	0.4108	0.938	0.062
	DRP_{GR}	$W_{without}$	0.4048	0.913	0.087	0.4424	1.002	0.002
		W_{GR}	0.4201	1.018	0.018	0.4411	1.024	0.024
		ERC_P	0.4347	1.029	0.029	0.4497	1.046	0.046
	DRP_{VR}	$W_{without}$	0.2395	0.814	0.186	0.2538	0.760	0.240
		W_{GR}	0.2589	0.838	0.162	0.2589	0.815	0.185
		ERC_P	0.2624	0.789	0.211	0.2655	0.840	0.160
Fat	EBV	$W_{without}$	0.5180	0.992	0.008	0.5639	0.938	0.062
		$W_{classic}$	0.5140	0.922	0.078	0.5639	0.938	0.062
	DYD	$W_{without}$	0.4691	0.983	0.017	0.4852	0.974	0.026
		$W_{classic}$	0.4963	1.098	0.098	0.4919	1.002	0.002
	DRP_{GR}	$W_{without}$	0.5171	1.143	0.143	0.4914	1.001	0.001
		W_{GR}	0.5373	1.264	0.264	0.4933	1.025	0.025
		ERC_P	0.5357	1.248	0.248	0.4908	1.022	0.022
	DRP_{VR}	$W_{without}$	0.3971	1.088	0.088	0.3998	0.961	0.039
		W_{GR}	0.3974	1.066	0.066	0.4049	1.023	0.023
		ERC_P	0.3936	1.017	0.017	0.4025	1.007	0.007
Protein	EBV	$W_{without}$	0.3648	0.750	0.250	0.3890	0.778	0.222
		$W_{classic}$	0.3595	0.718	0.282	0.3890	0.778	0.222
	DYD	$W_{without}$	0.3107	0.807	0.193	0.3107	0.807	0.193
		$W_{classic}$	0.3153	0.824	0.176	0.3157	0.826	0.174
	DRP_{GR}	$W_{without}$	0.3263	0.817	0.183	0.3646	0.892	0.108
		W_{GR}	0.3512	0.926	0.074	0.3703	0.923	0.077
		ERC_P	0.3631	0.943	0.057	0.3788	0.944	0.056
	DRP_{VR}	$W_{without}$	0.2290	0.722	0.278	0.2289	0.722	0.278
		W_{GR}	0.2485	0.766	0.234	0.2485	0.766	0.234
		ERC_P	0.2468	0.741	0.259	0.2468	0.741	0.259
	Mean		0.3786	0.9207	0.1440	0.3878	0.9003	0.1098

De-regression was based on VanRaden (DRP_{VR}), Garrick (DRP_{GR}), Estimated breeding value (EBV) and Daughter yield deviation (DYD); $W_{classic}$, W_{GR} and ERC_P represent three methods for weighting diagonal elements of incident matrix of residual error in estimation of genomic breeding values; $W_{without}$ shows the diagonal elements of incident matrix of residual error not weighted; for genomic evaluation MY, FY and PY, the optimal scaling factors were $\tau = 1$ and $\omega = 0.4$, 0.2 and 0.5, respectively for EBV as response variable. for MY, FY and PY, the optimal scaling factors were $\tau = 1$, 0.2 and 1.3 and $\omega = 0.6$, 0.6 and 0.1, respectively for DRP_{GR} as response variable. for MY, FY and PY, the optimal scaling factors were $\tau = 1$ and $\omega = 0.6$, 0.6 and 1, respectively for DRP_{VR} as response variable. For MY, FY and PY, the optimal scaling factors were $\tau = 1.5$, 1 and 1.5 and $\omega = 1$, 0.7 and 1, respectively for DYD as response variable.

reliability and Dev from un-weighted DRP_{GR} for MY were 0.44 and 0.002, respectively. In GBLUP, the reliability and Dev from un-weighted EBV for FY were 0.52 and 0.008, respectively. Moreover, in ssGBLUP, the reliability and Dev of DRP_{GR} in un-weighted analysis for FY were 0.49 and 0.001, respectively. In ssGBLUP, the reliability and Dev in weighted DRP_{GR} analysis for PY were 0.38 and 0.056, respectively (Table 2).

The estimated reliabilities of the response variables are different among the bulls. This variability is incorporated in **D** matrix which could result in more reliable predictions (Vandenplas and Gengler, 2015). Therefore, two weighting methods in **D** matrix (W_{GR} and ERC_P) for DRP_{GR} and DRP_{VR} response variables were compared. Also, the same weighting method ($W_{classic}$) was compared for EBV and DYD. *The* W_{GR} is based on heritability,

https://doi.org/10.1017/S0022029922000395 Published online by Cambridge University Press

reliability and portion of genetic variance not explained by markers (Garrick *et al.*, 2009). In this study, the value of c for W_{GR} was assumed to be 0.1 according to other studies (Song *et al.*, 2018). If the c is very close to zero, the reliability of genomic prediction is higher and the bias is lower (Song *et al.*, 2018).

In cases where the EBV or DRP_{VR} were used as response variables, the bias of prediction was the highest for MY and PY. These results agree with reports from a simulation study on different de-regression methods (Calus *et al.*, 2016). The performance of response variables of DRP_{VR} and EBV are reported to be modest (Calus *et al.*, 2016). In this study, when DRP_{GR} was used as the response variable, the bias was lower compared with the DRP_{VR}. The results of another study showed the GEBV predicted from EBV as response variable was biased (Guo *et al.*, 2010). In a simulation study, it was concluded that the de-regression by DRP_{GR} was superior compared to DRP_{VR} method (Calus *et al.*, 2016). In the present study, the results show the bias of DYD is low, which is because of no double-counting in the analysis.

Comparison of ssGBLUP and GBLUP methods in genomic prediction

The effect of scaling factors in **H** when combining $\mathbf{G}^{-1}(\tau)$ and $\mathbf{A}_{22}^{-1}(\omega)$ on validation reliabilities was low, but the effect on bias was high. However, scaling factors were different for each response variable and trait. The optimal scaling factors for EBV and the three traits were $\tau = 1$ and $\omega = 0.4$, 0.2 and 0.5, respectively. For DRP_{GR} and the three traits, the optimal scaling factors were $\tau = 1$, 0.2 and 1.3 and $\omega = 0.6$, 0.6 and 0.1, respectively. For DRP_{VR} regardless of studied trait, the optimal scaling factors were $\tau = 1$ and $\omega = 0.6$, 0.6 and 1, respectively. For DYD, the optimal scaling factors were $\tau = 1.5$, 1 and 1.5 and $\omega = 1$, 0.7 and 1, respectively. The differences in τ and ω values for different response variables and traits are due to differences in formulations of response variables and the genetic architecture of traits. The ideal scaling factors are specific according to the population and trait (Oliveira *et al.*, 2019). The scaling factors which are estimated for milk, fat and protein of Iranian Holstein cattle can be used or referred for this population by other researchers.

In ssGBLUP, the average reliability of genomic predictions for the three traits was 0.39, which was 0.98% points higher than the average reliability from the GBLUP method. Moreover, the bias of predictions from ssGBLUP was lower than GBLUP. The average Dev for the three traits was 0.11 in ssGBLUP and 0.14 in GBLUP. The τ and ω parameters that are used for calculating **H** showed that ω reduced bias in genomic prediction (Tsuruta *et al.*, 2011). It is suggested that the optimal scaling factors decrease the possible inflation of genomic predictions (Misztal *et al.*, 2013). Using the optimal scaling factors in **H** matrix reduce bias and increase the reliability of prediction (Oliveira *et al.*, 2019).

In the present study the number of bulls in the reference population was relatively small (818 animals). The size of the reference population in other studies was 3,045 (Gao *et al.*, 2012) and 5,160 bulls (Song *et al.*, 2018). The accuracy of genomic evaluation depends on heritability of the trait, the method of prediction and the number of animals in the reference population (Goddard, 2009). Interestingly, the high reliability of predictions from ssGBLUP indicates the method can be used for predictions in populations with a small number of genotyped animals.

In the routine procedure of multi-step genomic evaluation, the EBV, de-regressed EBV, direct genomic value (DGV) and finally the GEBV is predicted. Also, in the multi-step, the **G** matrix is used for prediction of DGV. In the present study, multi-step method was used but instead of **G** matrix, the **H** matrix was used for predictions. The results show that using the **H** matrix increased the reliability and reduced the bias.

In conclusion, the type of response variable and weighting or unweighting the residuals affects the prediction performance of statistical methods. In ssGBLUP, the un-weighted DRP_{GR} as the response variable for MY and FY and weighted DRP_{GR} for PY outperformed other response variables. Generally, the ssGBLUP method outperformed the GBLUP method both in terms of reliability as well as bias.

Acknowledgements. The authors gratefully acknowledge the assistance of Animal Breeding Center of Iran for providing the data. We also thank Dr Guosheng Su from the University of Aarhus for helpful comments.

References

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

Calus M, Vandenplas J, Ten Napel J and Veerkamp R (2016) Validation of simultaneous deregression of cow and bull breeding values and derivation of appropriate weights. *Journal of Dairy Science* **99**, 6403–6419.

Christensen OF and Lund MS (2010) Genomic prediction when some animals are not genotyped. *Genetics Selection Evolution* **42**, 1–8.

Forni S, Aguilar I and Misztal I (2011) Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information. *Genetics Selection Evolution* **43**, 1–7.

Gao H, Christensen OF, Madsen P, Nielsen US, Zhang Y, Lund MS and Su G (2012) Comparison on genomic predictions using three GBLUP methods and two single-step blending methods in the Nordic Holstein population. *Genetics Selection Evolution* **44**, 1–8.

Garrick DJ, Taylor JF and Fernando RL (2009) Deregressing estimated breeding values and weighting information for genomic regression analyses. *Genetics Selection Evolution* **41**, 1–8.

Goddard M (1985) A method of comparing sires evaluated in different countries. *Livestock Production Science* **13**, 321–331.

Goddard M (2009) Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica* **136**, 245–257.

Guo G, Lund M, Zhang Y and Su G (2010) Comparison between genomic predictions using daughter yield deviation and conventional estimated breeding value as response variables. *Journal of Animal Breeding and Genetics* **127**, 423–432.

Hayes BJ, Bowman PJ, Chamberlain AJ and Goddard ME (2009) Invited review: genomic selection in dairy cattle: progress and challenges. *Journal of Dairy Science* **92**, 433–443.

Jairath L, Dekkers J, Schaeffer L, Liu Z, Burnside E and Kolstad B (1998) Genetic evaluation for herd life in Canada. *Journal of Dairy Science* **81**, 550–562.

Koivula M, Strandén I, Su G and Mäntysaari EA (2012) Different methods to calculate genomic predictions—comparisons of BLUP at the single nucleotide polymorphism level (SNP-BLUP), BLUP at the individual level (G-BLUP), and the one-step approach (H-BLUP). *Journal of Dairy Science* **95**, 4065–4073.

Legarra A, Aguilar I and Misztal I (2009) A relationship matrix including full pedigree and genomic information. *Journal of Dairy Science* **92**, 4656–4663.

Misztal I (2017) Studies on inflation of GEBV in single-step GBLUP for type. *interbull bulletin* 51.

Misztal I, Tsuruta S, Strabel T, Auvray B, Druet T and Lee D (2002) BLUPF90 and related programs (BGF90). In Proceedings of the 7th world congress on genetics applied to livestock production, pp. 743–744.

Misztal I, Aguilar I, Johnson D, Legarra A, Tsuruta S and Lawlor T (2009) A unified approach to utilize phenotypic, full pedigree and genomic

https://doi.org/10.1017/S0022029922000395 Published online by Cambridge University Press

information for a genetic evaluation of Holstein final score. *Interbull Bulletin* **40**, 240–240.

Misztal I, Tsuruta S, Aguilar I, Legarra A, VanRaden P and Lawlor T (2013) Methods to approximate reliabilities in single-step genomic evaluation. *Journal of Dairy Science* **96**, 647–654.

Misztal I, Lourenco D and Legarra A (2020) Current status of genomic evaluation. *Journal of Animal Science* **98**, 101–115.

Mrode RA (2014) *Linear Models for the Prediction of Animal Breeding Values*. CABI, Wallingford, UK.

Oliveira HR, Silva FF, Brito L, Rocha Guarini A, Jamrozik J and Schenkel F (2018) Comparing deregression methods for genomic prediction of test-day traits in dairy cattle. *Journal of Animal Breeding and Genetics* **135**, 97–106.

Oliveira HR, Lourenco DAL, Masuda Y, Misztal I, Tsuruta S, Jamrozik J, Brito LF, Silva FF and Schenkel FS (2019) Application of single-step genomic evaluation using multiple-trait random regression test-day models in dairy cattle. *Journal of Dairy Science* **102**, 2365–2377.

Sargolzaei M, Chesnais JP and Schenkel FS (2014) A new approach for efficient genotype imputation using information from relatives. *BMC Genomics* **15**, 1–12.

Song H, Li L, Zhang Q, Zhang S and Ding X (2018) Accuracy and bias of genomic prediction with different de-regression methods. *Animal: An International Journal of Animal Bioscience* **12**, 1111–1117.

Tsuruta S, Misztal I, Aguilar I and Lawlor T (2011) Multiple-trait genomic evaluation of linear type traits using genomic and phenotypic data in US holsteins. *Journal of Dairy Science* **94**, 4198–4204.

Vandenplas J and Gengler N (2015) Strategies for comparing and combining different genetic and genomic evaluations: a review. *Livestock Science* **181**, 121–130.

VanRaden PM (2008) Efficient methods to compute genomic predictions. *Journal of Dairy Science* **91**, 4414–4423.

VanRaden P and Wiggans G (1991) Derivation, calculation, and use of national animal model information. *Journal of Dairy Science* **74**, 2737–2746.

VanRaden P, Van Tassell C, Wiggans G, Sonstegard T, Schnabel R, Taylor J and Schenkel F (2009) Invited review: reliability of genomic predictions for North American Holstein bulls. *Journal of Dairy Science* **92**, 16–24.

Vitezica Z-G, Aguilar I, Misztal I and Legarra A (2011) Bias in genomic predictions for populations under selection. *Genetics Research* **93**, 357–366.

https://doi.org/10.1017/S0022029922000395 Published online by Cambridge University Press