Open Access
Asian-Australas J Anim Sci
Vol. 32, No. 2:159-169 February 2019
https://doi.org/10.5713/ajas.18.0161
pISSN 1011-2367 eISSN 1976-5517

AJAS
Asian-Australasian Journal of Animal Sciences

Effects of preselection of genotyped animals on reliability and bias of genomic prediction in dairy cattle

Kenji Togashi[1,]*, Kazunori Adachi[2], Kazuhito Kurogi[1], Takanori Yasumori[2], Kouichi Tokunaka[2], Atsushi Ogino[1], Yoshiyuki Miyazaki[1], Toshio Watanabe[1], Tsutomu Takahashi[2], and Kimihiro Moribe[2]

* **Corresponding Author:** Kenji Togashi
Tel: +81-0272692440, **Fax:** +81-0272692440,
E-mail: k-togashi@liaj.or.jp

[1] Maebashi Institute of Animal Science, Livestock Improvement Association of Japan, Maebashi, Gunma 371-0121, Japan
[2] Livestock Improvement Association of Japan, Koto-ku, Tokyo 135-0041, Japan

ORCID
Kenji Togashi
https://orcid.org/0000-0002-5392-2493
Kazunori Adachi
https://orcid.org/0000-0002-5961-1792
Kazuhito Kurogi
https://orcid.org/0000-0002-1645-9196
Takanori Yasumori
https://orcid.org/0000-0003-3272-2254
Kouichi Tokunaka
https://orcid.org/0000-0002-2794-3223
Atsushi Ogino
https://orcid.org/0000-0002-6971-8745
Yoshiyuki Miyazaki
https://orcid.org/0000-0003-3682-2551
Toshio Watanabe
https://orcid.org/0000-0001-8827-5196
Tsutomu Takahashi
https://orcid.org/0000-0002-2074-8484
Kimihiro Moribe
https://orcid.org/0000-0003-1285-6060

Submitted Mar 1, 2018; Revised Apr 29, 2018;
Accepted Jun 22, 2018

Objective: Models for genomic selection assume that the reference population is an unselected population. However, in practice, genotyped individuals, such as progeny-tested bulls, are highly selected, and the reference population is created after preselection. In dairy cattle, the intensity of selection is higher in males than in females, suggesting that cows can be added to the reference population with less bias and loss of accuracy. The objective is to develop formulas applied to any genomic prediction studies or practice with preselected animals as reference population.
Methods: We developed formulas for calculating the reliability and bias of genomically enhanced breeding values (GEBV) in the reference population where individuals are preselected on estimated breeding values. Based on the formulas presented, deterministic simulation was conducted by varying heritability, preselection percentage, and the reference population size.
Results: The number of bulls equal to a cow regarding the reliability of GEBV was expressed through a simple formula for the reference population consisting of preselected animals. The bull population was vastly superior to the cow population regarding the reliability of GEBV for low-heritability traits. However, the superiority of reliability from the bull reference population over the cow population decreased as heritability increased. Bias was greater for bulls than cows. Bias and reduction in reliability of GEBV due to preselection was alleviated by expanding reference population.
Conclusion: Cows are easier in expanding reference population size compared with bulls and alleviate bias and reduction in reliability of GEBV of bulls which are highly preselected than cows by expanding the cow reference population.

Keywords: Reliability of Selection; Genomic Selection; Reference Population; Dairy Cattle

INTRODUCTION

Genomic prediction (GP) is used to predict the genomic breeding values of genotyped individuals [1]. The GP models usually do not account for selection. However, the reference population which is used for estimating marker effects with GP models usually consisted of progeny test bulls which was highly selected. Therefore, the prediction models are unable to incorporate past selection based on pedigree and phenotypes, perhaps leading to bias as well as decreased accuracy.

A formula for approximating the reliability and bias of the genomically enhanced breeding values (GEBV) that accounted for the prior selection of genotyped test bulls from among all test bull candidates was proposed [2]. In that method, the differences between the means and standard deviations of the estimated breeding values (EBV) of all of the test bull candidates are used to estimate the proportion of selective genotyping. Then, the selection difference or intensity of selection is calculated from quantitative genetics textbooks [3], and the authors approximated the reliability and bias of the GEBV by accounting for the effect of the intensity of selection [2]. However, the true genetic variance was reduced not only by the intensity of selection but also by the reliability of EBV [3]. The reliability of EBV differs by trait and between males and females. In dairy cattle populations, the intensity of selection is higher in males than in females [4], suggesting that cows potentially could be added to the reference population with less bias and loss of accuracy.

The genotyping of cows has become more prevalent as the cost of genotyping, in general, has decreased. The same individuals should be both genotyped and phenotyped, instead of genotyping the parents and phenotyping their progeny [5]. Adding genotyped females and their phenotypic records to the existing sire reference population is expected to increase the reliability of GPs, and increase in reliability would lead to increase in genetic gain and decreasing inbreeding in dairy cattle [6].

The genetic correlation between phenotypes of bulls and cows was approximately 0.6 for all yield traits and differed significantly from 1 [7]. Using selection index theory, the reliability of GEBV for a reference population in which the information contents in their phenotypes differed between groups, i.e., a reference population consisting of sires and cows both was presented [8]. However, in that method, the effect of preselection on reliability was not taken into account. Revaluation of cows would be beneficial from the standpoints of their less intense selection and easier incorporation into the reference population for its expansion, especially from the standpoints based on the bias and accuracy of GPs by varying the magnitudes of heritability and intensity of preselection among the animals chosen to create the reference population. A deterministic prediction model is necessary to develop simple formulas for calculating the reliability and bias of GEBV that accounts for prior selection of animals in the reference population. There are some results about using cows in the reference population with real data [9-12]. Construction of the real mixed reference population based on the deterministic model presented would improve the reliability of GEBV.

The first objective of the current study was to develop formulas for calculating the reliability and bias of GPs in which the effects of both intensity of selection and the reliability of EBV of preselected animals in the reference population would be taken into account. Next is to present a formula to calculate the number of bulls equal to a cow in regard to creating the same reliabilities of the GEBV between preselected bulls and cows in the reference population. The last is to present a guideline to create a reference population composed of preselected bulls and cows to prevent the reduction of reliability and bias of GEBV due to preselection before the actual creation of the reference population.

MATERIALS AND METHODS

Formulas for reliability and bias under selection

The variance of true breeding value of the animals selected on the basis of EBV can be expressed as:

$$\sigma^2_{G*} = \sigma^2_G(1 - kr^2_{EBV}),$$

where * denotes values after selection, G is true breeding value, σ^2_G is the variance of G before selection, k is the proportional reduction in the variance of the selection criterion due to selection, and r^2_{EBV} is the reliability of EBV. With truncation selection on a normally distributed selection criterion, k is determined entirely by the intensity of selection, k = i (i – x), where i is the intensity of selection, and x is the standardized truncation point [3]. Similarly, the variance of the GEBV of the animals selected on the basis of EBV can be expressed as:

$$\sigma^2_{GEBV*} = \sigma^2_{GEBV}(1 - kr^2_{EBV,GEBV}),$$

where σ^2_{GEBV} is the variance of the GEBV before selection, and $r^2_{EBV,GEBV}$ is the squared correlation between EBV and GEBV.

Assuming prediction error cov(EBV, GEBV) is zero, the correlation between EBV and GEBV can be shown as:

$$r_{EBV,GEBV} = r_{EBV}r_{GEBV}.$$

Therefore, the variance of the GEBV after selection is:

$$\sigma^2_{GEBV*} = \sigma^2_{GEBV}(1 - kr^2_{EBV,GEBV}) = \sigma^2_{GEBV}(1 - kr^2_{EBV}r^2_{GEBV}).$$

In addition, the cov(EBV, GEBV) can be written as:

$$\sigma_{EBV,GEBV} = r_{EBV}r_{GEBV}\sigma_{EBV}\sigma_{GEBV} = r^2_{EBV}r^2_{GEBV}\sigma^2_G.$$

A general expression for a covariance after selection [13] is:

$$\sigma^*_{jk} = \sigma_{jk} - k\frac{\sigma_{ij}\sigma_{ik}}{\sigma^2_i},$$

where σ^*_{jk} is the covariance between j and k selected on i, and σ_{jk} is the covariance before selection. Therefore the cov(GEBV, G) after selection on EBV is:

$$\begin{aligned}\sigma_{G,GEBV^*} &= \sigma_{G,GEBV} - k\frac{\sigma_{G,EBV}\sigma_{GEBV,EBV}}{\sigma^2_{EBV}}\\ &= \sigma^2_{GEBV} - k\sigma_{GEBV,EBV} = \sigma^2_{GEBV} - kr^2_{EBV}r^2_{GEBV}\sigma^2_G\\ &= \sigma^2_{GEBV}(1 - kr^2_{EBV}).\end{aligned}$$

In summary, the reliability of GEBV after accounting for selection on EBV can be expressed based on the reliability of GEBV under random selection (r^2_{GEBV}), the intensity of selection, and the reliability of EBV (r^2_{EBV}) of the preselected individuals both genotyped and phenotyped in the reference population:

$$r^2_{GEBV,G^*} = \frac{\sigma^2_{G,GEBV^*}}{\sigma^2_{G^*}\sigma^2_{GEBV^*}} = \frac{\sigma^2_{GEBV}(1-kr^2_{EBV})}{\sigma^2_{G}(1-kr^2_{EBV}r^2_{GEBV})}$$
$$= r^2_{GEBV}\frac{1-kr^2_{EBV}}{1-kr^2_{EBV}r^2_{GEBV}} \qquad (1)$$

When $r^2_{EBV} = 1$, equation (1) yields the same formula as that in [2].

The regression coefficient of G or deregressed proofs on GEBV is a criterion for bias in GEBV [2,14,15]. Accordingly, the regression coefficient of G on GEBV after selection by using EBV is written as:

$$b^* = \frac{\sigma_{G,GEBV^*}}{\sigma^2_{GEBV^*}} = \frac{\sigma^2_{GEBV}(1-kr^2_{EBV})}{\sigma^2_{GEBV}(1-kr^2_{EBV}r^2_{GEBV})}$$
$$= \frac{1-kr^2_{EBV}}{1-kr^2_{EBV}r^2_{GEBV}} \qquad (2)$$

When $r^2_{EBV} = 1$, equation (2) results in the same regression coefficient as in [2]. That is, we extended the formula for the reliability and bias of GEBV from [2] by accounting for the preselection on EBV of the animals used to create the reference population.

Estimating the number of bulls equal to a cow in the reliability of the GEBV of preselected animals in the reference population

Reliability in a reference population after selection was expressed as (1). The reliability of GEBV without selection ($r^2_{GEBV,G}$) or under random selection is obtained after transformation of (1):

$$r^2_{GEBV,G} = \frac{r^2_{GEBV,G^*}}{1-kr^2_{EBV}+kr^2_{GEBV,G^*}r^2_{EBV}} \qquad (3)$$

The reliabilities of GEBVs depend on the size of the reference population (nP), the effective number of loci for which effects have to be estimated (nG), and the correlation of the G of a genotyped individual with its phenotypic record (r). In a random sample of the population, the reliability of GEBV or the correlation between GEBV and G ($r^2_{GEBV,G}$) can be calculated as described in [16]:

$$r^2_{GEBV,G} = \frac{\lambda r^2}{\lambda r^2 + 1} \qquad (4)$$

where λ = nP/nG. Parameter nG depends on the historical effective size of the unselected population (N_E) and on the size of the genome, L (in Morgans), and can be estimated as shown in [17]:

$$nG = 2N_E L$$

When an individual in the reference population is both genotyped and phenotyped, r is equal to the square root of heritability of the trait. Then, the reliability of cows from their own records is:

$$r^2 = h^2$$

When the reference population is based on progeny-tested sires, i.e., when sires are genotyped but their offspring are phenotyped, r equals the accuracy of the EBV obtained from progeny testing [5]:

$$r^2 = \frac{0.25Nh^2}{1+0.25(N-1)h^2},$$

where N is the number of half-sibling progeny on which the EBV is based.

Parameter nP can be transformed from (4):

$$nP = nG\frac{r^2_{GEBV,G}}{r^2(1-r^2_{GEBV,G})} \qquad (5)$$

When the reference population is composed of either bulls (m) or cows (f), parameter nP in (5) can be written as:

$$nP_m = nG\frac{r^2_{m\_GEBV,G}}{r^2_m(1-r^2_{m\_GEBV,G})} \text{ or}$$

$$nP_f = nG\frac{r^2_{f\_GEBV,G}}{r^2_f(1-r^2_{f\_GEBV,G})},$$

where the subscript letters m and f refer to male and female, respectively.

Therefore,

$$\frac{nP_m}{nP_f} = \frac{r^2_f(1-r^2_{f\_GEBV,G})r^2_{m\_GEBV,G}}{r^2_m(1-r^2_{m\_GEBV,G})r^2_{f\_GEBV,G}} \qquad (6)$$

Alternatively, the reliability of GEBV under random selection can be written as (3) by using that of GEBV after preselection, therefore using the subscript letters m and f as defined earlier,

$$r^2_{m\_GEBV,G} = \frac{r^2_{m\_GEBV,G*}}{1 - k_m r^2_{m\_EBV} + k_m r^2_{m\_GEBV,G*} r^2_{m\_EBV}}$$

and

$$r^2_{f\_GEBV,G} = \frac{r^2_{f\_GEBV,G*}}{1 - k_f r^2_{f\_EBV} + k_f r^2_{f\_GEBV,G*} r^2_{f\_EBV}} \quad (7)$$

Substituting (7) into (6) yields:

$$\frac{nP_m}{nP_f} = \frac{r^2_f}{r^2_m} \times \frac{r^2_{m\_GEBV,G*}}{r^2_{f\_GEBV,G*}} \times \frac{(1 - k_f r^2_{f\_EBV})(1 - r^2_{f\_GEBV,G*})}{(1 - k_m r^2_{m\_EBV})(1 - r^2_{m\_GEBV,G*})} \quad (8)$$

The numbers of bulls equal to a cow in regard to the specific reliabilities of the GEBV of animals in the reference population with and without preselection are calculated by using (8) and (6), respectively. Note that $r^2_{f\_EBV}$ and $r^2_{m\_EBV}$ are the reliabilities of EBV for preselection used to create the females-only and males-only reference population, respectively. The number of bulls equal to a cow in terms of the reliability of GEBV must be compared under the same reliability of GEBV after preselection, i.e., $r^2_{m\_GEBV,G*} = r^2_{f\_GEBV,G*}$.

Therefore, the number of bulls equal to a cow in a standpoint of bringing about the same size of reliabilities of the GEBV of preselected animals in the reference population is:

$$\frac{nP_m}{nP_f} = \frac{r^2_f}{r^2_m} \times \frac{(1 - k_f r^2_{f\_EBV})}{(1 - k_m r^2_{m\_EBV})} \quad (9)$$

Note that the number of bulls equal to a cow in terms of the reliability of GEBV without preselection, i.e., k = 0, depends only on the reliability of EBV of the individuals both genotyped and phenotyped in the reference population.

Reliability of GEBV in the reference population consisting of preselected bulls and cows

Using selection index theory, the reliability of GEBV was derived by [10], which is explained by markers, in a reference population consisting of multiple groups of animals whose phenotypes differ in their information content. We extended selection index theory to a reference population consisting of preselected bulls and cows.

From selection index theory,

$$\begin{bmatrix} r^2_{m\_GEBV,G^*} & r^2_{m\_GEBV,G^*} r^2_{f\_GEBV,G^*} \\ r^2_{f\_GEBV,G^*} r^2_{m\_GEBV,G^*} & r^2_{f\_GEBV,G^*} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} r^2_{m\_GEBV,G^*} \\ r^2_{f\_GEBV,G^*} \end{bmatrix}$$

$$r^2_{m+f^*} = b_1 r^2_{m\_GEBV,G^*} + b_2 r^2_{f\_GEBV,G^*} \quad (10)$$

where r^2_{m+f*} is the reliability of GEBV in the reference population consisting of preselected bulls and cows.

The increase in reliability from including bulls only to both bulls and cows in a reference population is expressed as the difference between the reliability of GEBV in the reference population consisting of preselected bulls and cows and that of including preselected bulls only, i.e., $r^2_{m+f*} - r^2_{m\_GEBV,G*}$. This increase in reliability corresponds to the increase after preselection and therefore can be converted to the increase under random selection. That is, the increase ($r^2_{\Delta\_m+f}$) in the reliability of GEBV under random selection due to the addition of cows into the reference population relative to the reference population containing males only can be expressed applying (3):

$$r^2_{\Delta\_m+f} = \frac{r^2_{m+f*} - r^2_{m\_GEBV,G*}}{1 - k_m r^2_{m\_EBV} + k_m r^2_{m\_EBV}(r^2_{m+f*} - r^2_{m\_GEBV,G*})} \quad (11)$$

The number of bulls corresponding to this increase ($nP_{m-\Delta}$) can be computed applying (5):

$$nP_{m-\Delta} = nG \frac{r^2_{\Delta\_m+f}}{r^2_m (1 - r^2_{\Delta\_m+f})}$$

We designated the number of cows in the reference population as nP_f. The increase in reliability is derived from adding cows into the reference population that consists of bulls only. That is, the number of bulls equal to a cow in regard to the reliability of GEBV in the reference population consisting of preselected bulls and cows (Cow_{value_m+f}) is:

$$Cow_{value\_m+f} = \frac{nP_{m-\Delta}}{nP_f} = \frac{nG \times r^2_{\Delta\_m+f}}{nP_f \times r^2_m (1 - r^2_{\Delta\_m+f})} \quad (12)$$

Simulation data

We preselected animals in the reference population according to the EBV of the trait of interest rather than selecting them randomly, to obtain more realistic reference populations [14,18,19]. The animals in the reference population came from several generations in the past but were approximated and simplified to come from a single generation, i.e., they were preselected on EBV, and the reference population was created from the phenotypic data of the preselected bulls' daughters' records or the preselected cows' own data. When the reference population is based on progeny-tested bulls, the number of daughters per test bull was set to 50 and 100. All test bull candidates were assumed to be preselected according to the PA (parent average). PA was computed by using the EBVs of sire (from 50 and 100 daughters) and dam. When the number of daughters per progeny-tested bull was set to 50, PA was calculated from EBV of sire from 50 daughters. That is, same number was set to the number of daughters of sire in PA and that of progeny-tested bull. Note that bulls for progeny testing were preselected from all test bull candidates, and they became test bulls after preselection and progeny-tested sires with their daughters' records (50 or 100) after progeny-testing. Heifers were preselected using their PA, and cows were preselected according to the EBV from their own records. After preselection, test bulls, heifers, and cows were used to create the reference population. The reliability of the EBV from their own records was calculated by selection index theory where reliabilities of PA and their individual records constituted the index similar to the equation (10) and the number of daughters of sire in PA was set to 50.

We assumed that the length of the genome was 30 Morgans and that the heritability of the trait of interest was 0.1, 0.3, or 0.5. The historical effective population size was set to 100 animals [5,20]. The preselection percentage on EBV of animals used to create the reference population was set to 5%, 30%, and 100% for males and to 70%, 90%, and 100% for females. When animals were selected randomly, the proportional reduction (k) in the variance of G was set to zero. The reference population size was set to 5,000, 10,000, 20,000, and 40,000.

RESULTS

Reliability of GEBV of preselected animals in the reference population

We calculated the reliability of the GEBV of non-preselected animals and the ratio of the reliability of preselected animals to that of non-preselected animals for reference populations composed solely of proven bulls preselected on PA computed by using EBVs of sire (from 50 or 100 daughters) and dam (Table 1). The reliability of the GEBV of cows was shown in Table 2. The reliability of preselection on PA, EBV from a cow's own record, or a bull's progeny testing based on 50 daughters was calculated at three levels of heritability (Table 3). The reliability of GEBV in the bulls-only reference population was the highest among the three reference populations (bulls preselected on PA, heifers preselected on PA, and cows preselected on EBV from their own records). The bulls-only population was particularly superior to the cow population for low-heritability traits ($h^2 = 0.1$), especially for the bulls-only population testing based on 100 daughters. However, the superiority of the reliability associated with the bull reference population decreased as heritability increased, regardless of whether the animals in the reference population were preselected or not.

In addition, the reliability of GEBV decreased as the intensity of preselection increased (i.e., a decrease in the preselection percentage), and this trend became more conspicuous as heritability increased. This change occurs because the effect of preselection on the reduction of the variance of G increases as heritability increases. The decrease in the reliability of preselected animals compared with that of non-preselected animals became more conspicuous as the reference population size decreased. That is, the effect of preselection on the decrease in reliability became more deleterious as the reference population became smaller.

Bias of GEBV

Regression coefficients of G on GEBV for animals in the reference populations composed solely of proven bulls preselected on PA calculated by using EBVs of sire (from 50 and 100 daughters) and dam were shown (Table 4). The regression

Table 1. The reliability of GEBV of non-preselected bulls at 100% preselection and the ratio of reliability of preselected bulls at <100% preselection to that of non-preselected bulls

Heritability	No. of animals	Bulls preselected according to PA[1] (progeny testing 50 daughters per test bull)			Bulls preselected according to PA[2] (progeny testing 100 daughters per test bull)		
		Preselection percentage (%)					
		5	30	100	5	30	100
0.1	5,000	0.8982	0.9137	0.3189	0.8817	0.9001	0.3748
	10,000	0.9209	0.9332	0.4837	0.9111	0.9253	0.5453
	20,000	0.9452	0.9540	0.6519	0.9406	0.9503	0.7057
	40,000	0.9661	0.9716	0.7893	0.9643	0.9703	0.8275
0.3	5,000	0.8426	0.8677	0.4006	0.8346	0.8613	0.4259
	10,000	0.8823	0.9018	0.5721	0.8780	0.8986	0.5974
	20,000	0.9218	0.9352	0.7278	0.9200	0.9340	0.7479
	40,000	0.9532	0.9615	0.8425	0.9526	0.9611	0.8558
0.5	5,000	0.8039	0.8361	0.4223	0.7989	0.8322	0.4378
	10,000	0.8536	0.8789	0.5938	0.8510	0.8770	0.6090
	20,000	0.9028	0.9204	0.7452	0.9019	0.9198	0.7570
	40,000	0.9419	0.9528	0.8540	0.9417	0.9527	0.8617

GEBV, genomically enhanced breeding value; PA, parental average.

[1] PA was calculated by using EBVs from sire (from 50 daughters) and dam.

[2] PA was calculated by using EBVs from sire (from 100 daughters) and dam.

Table 2. The reliability of GEBV of cows

Heritability	No. of animals	Cows preselected according to PA[1]				Cows preselected according to the EBV from their individual records			
		Preselection percentage (%)							
		70	80	90	100	70	80	90	100
0.1	5,000	0.0683	0.0699	0.0721	0.0769	0.0709	0.0720	0.0735	0.0769
	10,000	0.1279	0.1306	0.1344	0.1429	0.1325	0.1343	0.1370	0.1429
	20,000	0.2269	0.2311	0.2370	0.2500	0.2339	0.2368	0.2410	0.2500
	40,000	0.3698	0.3754	0.3832	0.4000	0.3792	0.3830	0.3883	0.4000
0.3	5,000	0.1620	0.1690	0.1789	0.2000	0.1770	0.1812	0.1871	0.2000
	10,000	0.2788	0.2891	0.3034	0.3333	0.3008	0.3068	0.3152	0.3333
	20,000	0.4361	0.4486	0.4656	0.5000	0.4625	0.4695	0.4793	0.5000
	40,000	0.6073	0.6193	0.6354	0.6667	0.6324	0.6390	0.6481	0.6667
0.5	5,000	0.2243	0.2376	0.2561	0.2941	0.2559	0.2630	0.2729	0.2941
	10,000	0.3664	0.3840	0.4077	0.4546	0.4075	0.4164	0.4288	0.4546
	20,000	0.5363	0.5549	0.5793	0.6250	0.5791	0.5880	0.6002	0.6250
	40,000	0.6981	0.7138	0.7336	0.7692	0.7334	0.7406	0.7502	0.7692

GEBV, genomically enhanced breeding valuel; PA, parental average; EBV, estimated breeding value.

[1] PA was calculated by using EBVs from sire (from 50 daughters) and dam.

Table 3. Reliability of a bull's progeny testing, cow's estimated breeding value (EBV) based on her individual record and parental average (PA[1]), and preselection on PA[1]

Heritability	Bull's progeny testing[2]	Cow's EBV	Preselection on PA
0.1	0.562 (1.0) 3)	0.236 (1.0)	0.165 (1.0)
0.3	0.802 (1.428)	0.447 (1.893)	0.276 (1.666)
0.5	0.877 (1.561)	0.604 (2.556)	0.344 (2.082)

[1] PA was calculated by using EBVs from sire (from 50 daughters) and dam.

[2] 50 daughters per test bull.

[3] The figures within parentheses are the ratio of reliability to that for a heritability of 0.1.

coefficients of G on GEBV for cows preselected on EBV were calculated (Table 5). Regression coefficients of G on GEBV deviated more from 1 as the intensity of preselection increased, thus indicating that overestimation of GEBV became more prominent as the intensity of preselection increased. Because the intensity of preselection is higher in bulls than in cows, bias was more problematic in bulls than in cows. When the cow reference population preselected on PA was compared with that preselected by using the EBV from their own records, bias or overestimation of GEBV was greater for cows preselected on the EBV from their own records than those

Table 4. Regression coefficients of true breeding values on the GEBV of preselected bulls in the reference population[1]

Heritability	No. of animals	Bulls preselected according to PA[2] (progeny-testing 50 daughters per test bull)		Bulls preselected according to PA[3] (progeny-testing 100 daughters per test bull)	
		Preselection percentage (%)			
		5	30	5	30
0.1	5,000	0.898	0.914	0.882	0.900
	10,000	0.921	0.933	0.911	0.925
	20,000	0.945	0.954	0.941	0.950
	40,000	0.966	0.972	0.964	0.970
0.3	5,000	0.843	0.868	0.835	0.861
	10,000	0.882	0.902	0.878	0.899
	20,000	0.922	0.935	0.920	0.934
	40,000	0.953	0.961	0.953	0.961
0.5	5,000	0.804	0.836	0.799	0.832
	10,000	0.854	0.879	0.851	0.877
	20,000	0.903	0.920	0.902	0.920
	40,000	0.942	0.953	0.942	0.953

GEBV, genomically enhanced breeding valuel; PA, parental average; EBV, estimated breeding value.

[1] In all cases when the preselection percentage was 100%, the regression coefficient was 1.0.

[2] PA was calculated by using EBVs from sire (from 50 daughters) and dam.

[3] PA was calculated by using EBVs from sire (from 100 daughters) and dam.

Table 5. Regression coefficients of true breeding values on the GEBV of preselected cows in the reference population[1)]

Heritability	No. of animals	Cows preselected according to the EBV of their individual records		Cows preselected according to PA[2)]	
		Preselection percentage (%)			
		70	90	70	90
0.1	5,000	0.888	0.937	0.922	0.956
	10,000	0.896	0.941	0.927	0.959
	20,000	0.907	0.948	0.936	0.964
	40,000	0.925	0.958	0.948	0.971
0.3	5,000	0.810	0.894	0.885	0.935
	10,000	0.837	0.910	0.902	0.946
	20,000	0.872	0.931	0.925	0.959
	40,000	0.911	0.953	0.949	0.972
0.5	5,000	0.762	0.871	0.870	0.928
	10,000	0.806	0.897	0.897	0.943
	20,000	0.858	0.927	0.927	0.960
	40,000	0.908	0.954	0.953	0.975

GEBV, genomically enhanced breeding valuel; EBV, estimated breeding value; PA, parental average.

[1)] In all cases when the preselection percentage was 100%, the regression coefficient was 1.0.

[2)] PA was calculated by using EBVs from sire (from 50 daughters) and dam.

preselected on PA because the reliability of EBV from their individual record was greater than that of PA (Table 3). In the same way, bias or overestimation of GEBV was greater for bulls testing 100 daughters than 50 daughters. That is, bias became more pronounced with an increase in the reliability of preselection of animals used to create the reference population. Bias or overestimation of GEBV was alleviated by increasing reference population size (Tables 4, 5).

The contribution to the same reliability of the number of bulls to a cow

The number of bulls equal to a cow in terms of the bringing about the same size of reliability of the GEBV of preselected animals was calculated by (9) (Table 6). This parameter is related solely to the reliability ($r^2_{f\_EBV}, r^2_{m\_EBV}$) of the preselection of heifers/cows and bulls to create the reference population, the intensity of preselection (k_f, k_m), and the reliability of the EBV of cow's record or bull's progeny testing (r^2_f, r^2_m). The number of bulls equal to a cow in regard to the reliability increased with increases in the intensity of preselection for males and with decreases in the intensity of selection for females. The number increased three to four times with an increase in heritability from 0.1 to 0.5 under the same preselection percentage. For example, the number of bulls equal to a cow increased approximately four times, from 0.208 to 0.811, as heritability increased from 0.1 to 0.5 under the 5% male preselection percentage and random female preselection.

Reliability of the GEBV in the reference population comprising both bulls and cows

The combined reliabilities in the reference population composed of 10,000 preselected bulls and 10,000 or 20,000 preselected cows are calculated by (10) and shown together with the reliabilities of reference populations composed solely of bulls or cows (Table 7). The cows in Table 7 are only those preselected on EBV from their individual records, because the combined reliability was almost equivalent whether heifers were preselected on PA or cows were preselected on the EBV from their own records. The combined reliability increased as the number of cows increased from 10,000 to 20,000, and this trend was more conspicuous for high-heritability traits ($h^2 = 0.5$) than low-heritability traits ($h^2 = 0.1$). As shown in Table 2 and 6, this result again confirmed cows' favorable properties regarding the reliability of high-heritability traits. The contribution of cows in reliability of the combined population compared with that of a reference population composed of bulls only, i.e., the difference between combined reliability due to bulls and cows and the reliability due to bulls only, ranged from 0.03 to 0.22. The reliability for a reference population composed of either bulls or cows solely was computed by using (1). The number of bulls equal to a cow in terms of the reliability of

Table 6. The number of bulls equal to a cow in regard to bringing about the same size of reliability of GEBV of preselected animals in the reference population

Heritability	Preselection (%) of bulls according to PA[1)]	Preselection (%) of cows according to the EBV of their individual records		Preselection (%) of cows according to PA[1)]	
		Preselection percentage of cows (%)			
		70	100	70	100
0.1	5	0.183	0.208	0.190	0.208
0.1	30	0.178	0.203	0.186	0.203
0.3	5	0.379	0.490	0.422	0.490
0.3	30	0.363	0.469	0.403	0.469
0.5	5	0.562	0.811	0.669	0.811
0.5	30	0.530	0.763	0.630	0.763

GEBV, genomically enhanced breeding valuel; PA, parental average; EBV, estimated breeding value.

[1)] PA was calculated by using EBVs from sire (from 50 daughters) and dam.

Table 7. Reliability of GEBV of the mixed reference population comprising preselected bulls and cows

Heritability	No. of cows	Preselection percentage (%) bulls - cows	Reliability of bulls only (progeny-testing 100 daughters per test bull)	Reliability of cows only	Reliability
0.1	10,000	5 - 70	0.497	0.126	0.531
	10,000	5 - 100	0.497	0.143	0.536
	10,000	30 - 70	0.505	0.126	0.538
	10,000	30 - 100	0.505	0.143	0.542
	20,000	5 - 70	0.497	0.223	0.560
	20,000	5 - 100	0.497	0.250	0.569
	20,000	30 - 70	0.505	0.223	0.566
	20,000	30 - 100	0.505	0.250	0.575
0.3	10,000	5 - 70	0.524	0.277	0.598
	10,000	5 - 100	0.524	0.333	0.616
	10,000	30 - 70	0.537	0.277	0.607
	10,000	30 - 100	0.537	0.333	0.624
	20,000	5 - 70	0.524	0.434	0.652
	20,000	5 - 100	0.524	0.500	0.670
	20,000	30 - 70	0.537	0.434	0.658
	20,000	30 - 100	0.537	0.500	0.683
0.5	10,000	5 - 70	0.518	0.365	0.623
	10,000	5 - 100	0.518	0.455	0.656
	10,000	30 - 70	0.534	0.365	0.633
	10,000	30 - 100	0.534	0.455	0.664
	20,000	5 - 70	0.518	0.535	0.690
	20,000	5 - 100	0.518	0.625	0.733
	20,000	30 - 70	0.534	0.535	0.697
	20,000	30 - 100	0.534	0.625	0.738

GEBV, genomically enhanced breeding valuel; EBV, estimated breeding value.
Cows were preselected according to the EBV of their individual records.
Bulls were preselected on PA calculated by using EBVs from sire (from 100 daughters) and dam.
No. of bulls in the mixed reference population was 10,000.

the reference population created from both preselected bulls and cows computed by using (12) coincided with the number computed by using (9). That is, the number of bulls equal to a cow in regard to the reliability of GEBV in the reference population comprising both bulls and cows agreed with the number of bulls equal to a cow in the reference population created solely from bulls or cows in Table 6.

DISCUSSION

Benefit of cows regarding the reliability of GEBV for high-heritability traits

The superiority of the reliability of the GEBV from a bulls-only reference population over the cow population decreased as heritability increased regardless of whether animals in the reference population were preselected (Tables 1, 2). To improve GP, the same individuals should be both genotyped and phenotyped instead of genotyping parents and phenotyping their progeny [5]. In the current study, cows are both genotyped and phenotyped, whereas bulls are genotyped and their progeny are phenotyped. Because the reliability of a cow's EBV was based on her own record, the increase in reliability concurrent with an increase in heritability was greater for cows than for bulls (Table 3). For example, the reliability of a cow's EBV based on her own record and that of a bull's EBV based on 50 of his daughters' records corresponding to a heritability of 0.1 are 0.236 and 0.562; when heritability is 0.5, these are 0.604 and 0.877, respectively. Consequently, we consider that genotyping of cows with phenotypes is advantageous for high-heritability traits from the point of increasing the reliability of GEBV. The value of genotyping of cows with phenotypes was reduced by increasing the number of daughters per test bull (results not shown), because the reliability of bulls increased with increases in the number of daughters per test bull (Table 1).

The effects of preselection on reliability and bias of GEBV

The effect of preselection on reducing the variance of G increased as heritability increased, thereby decreasing the reliability of the GEBV of preselected animals. However, the reliability of GEBV in the reference population increased as heritability increased even if preselection had been practiced (Tables 1, 2). This result indicates that the effect of the increase in heritability on the increase in the reliability of the EBV of animals by using a cow's own record or progeny testing of bulls was greater than that on the decrease in reliability due to reduction of the variance of G from preselection.

The effect of preselection on the decreased reliability of GEBV became more deleterious for smaller reference populations (Tables 1, 2). This result is explained by (1). That is, the reliability of GEBV after preselection is written as:

$$r^2_{GEBV,G*} = r^2_{GEBV} \frac{1 - kr^2_{EBV}}{1 - kr^2_{EBV} r^2_{GEBV}} ,$$

where r_{GEBV} is the accuracy of GEBV in the absence of preselection or under random selection. By contrast, the ratio of reliability after preselection ($r^2_{GEBV,G*}$) to that under random selection (r^2_{GEBV}) is written as:

$$\frac{r^2_{GEBV,G*}}{r^2_{GEBV}} = \frac{1 - kr^2_{EBV}}{1 - kr^2_{EBV} r^2_{GEBV}}$$

The ratio is 1.0 without loss of reliability when $r^2_{GEBV} = 1.0$. The ratio becomes smaller with decreases in reliability under random selection (r^2_{GEBV}) when the reliability of preselection (r^2_{EBV}) and the intensity of preselection are held constant. The main factor to decrease the reliability of GEBV under random selection (r^2_{GEBV}) is a reduction in the reference population size. Therefore, the effect of decreasing the reliability of GEBV due to preselection became much greater as the size of the reference population decreased.

Bias emerged when the cow reference population was under selection, but it decreased with increase in the size of the cow reference population (Table 5). The inclusion of cows in the reference population slightly reduced the bias in GEBV [17]. Routine genotyping of heifer calves or yearling heifers can be a cost-effective strategy for enhancing the genetic value of replacement females on commercial dairy farms [21]. That is, saying that a replacement decision was based on GEBV implies that genotyped heifer calves or yearling heifers were included in the reference population. In general, it is easier to increase reference population size by using cows than bulls. Expanding the reference population alleviated bias when heritability and intensity of preselection remained constant (Tables 4, 5). This effect occurs because the regression coefficient as a criterion of bias is defined as $(\frac{1 - kr^2_{EBV}}{1 - kr^2_{EBV} r^2_{GEBV}})$ as shown in (2) and because the reliability of GEBV (r^2_{GEBV}) under random selection increases with expanding reference population size. Note that the regression coefficient as a criterion of bias is equal to the ratio of reliability after preselection ($r^2_{GEBV,G*}$) to that under random selection ($r^2{}_{GEBV}$), i.e., . $\frac{r^2_{GEBV,G*}}{r^2_{GEBV}}$ That is, both of them can be written as: $\frac{1 - kr^2_{EBV}}{1 - kr^2_{EBV} r^2_{GEBV}}$. Consequently, bias and reduction in reliability of GEBV due to preselection was alleviated by expanding reference population. Therefore, cows can contribute to reducing bias and increasing reliability due to their ease of use in expanding reference population size and by providing more recent animals (compared with bulls). Cows' contribution was determined to improving GP in breeding schemes where few bulls with traditional evaluations were added annually [22,23]. Older bulls may contribute only slightly to increasing genomic reliability because of linkage decay between the validation and ancestral populations, resulting in r_g<1.0 between bulls and cows and lowering reliability [7]. The young selection candidates are more closely related to the animals in the reference population when the reference population consists of cows or a combination of bulls and cows instead of bulls only [7]. The GP is more reliable when juvenile animals share their recent pedigree with animals in the reference population [24,25]. Cows are easier in expanding reference population size compared with bulls and alleviate bias and reduction in reliability of bulls' GEBV due to higher preselection by expanding reference population of cows.

The value of cows compared with that of bulls in terms of the reliability of GEBV

The overall reliability in the reference population comprising both bulls and cows was not the sum of the reliabilities from those containing bulls or cows only (Table 7); this result indicated that marker information between bulls and cows was not independent, which was in agreement with the reliability results obtained from Danish cows and US bulls [26]. That is, the off-diagonal elements in (10) derived from index selection theory for the reference populations containing both bulls and cows were not zero.

The number of bulls equal to a cow in reliability of GEBV as calculated from (12) in the reference population containing both bulls and cows agreed with the number computed from (9) in a reference population created solely from bulls or cows. This effect occurs because the increased reliability due to the addition of cows into a bulls-only population was converted to the increase per head in the bulls-only population and the numbers of bulls only and cows only to yield the increased reliability was compared. Consequently, the number of bulls equal to a cow in terms of the reliability of the combined reference population could be computed by using the simple formula of (9) applied to reference populations created solely from bulls or cows. Cows are, in general, selected randomly compared with bulls; consequently, the effect of preselection on decreased reliability and bias of GEBV would be much smaller for cows than for bulls.

Assumption of parameters

The proportion of genetic variance explained by its markers is influenced by the effective size of the population (N_E) and the density at which the genetic analysis covers the genome. The number of independent segments present in the genome is expected to be lower at low (compared with high) N_E [27]. Therefore, the accuracy of GP is expected to be higher in a population with a smaller N_E than in a population with a larger N_E. An N_E of 750 was assumed by [8], whereas an N_E of 100 was assumed in the current study and by [5,20].

Simulation studies have shown that the accuracy of GEBV decreases slowly over generations when mating is random [1, 28] and more rapidly when selection is considered [29]. Given that recombination breaks up linkage disequilibrium (LD) in both situations, this finding indicates that selection is an important factor for decreasing LD between markers and qualitative trait loci. That is, accurate prediction of GEBV strongly depends on the persistence of LD between markers and qualitative trait loci across generations. However, we used a single generation in the current study to develop a simple formula for assessing the accuracy of GEBV that accounted for the effect of selection instead of accounting for persistent accuracy across generations. Advances in the use of sequence data and gene expression studies would lead to improved persistence of GP and potentially lead to greater reliabilities [12]. The development of methodology for estimating persistent accuracy of GEBV across generations is warranted.

An empirical value for the number of independent chromosome segments could be used in place of nG [6]. The accuracy of GEBV was also proposed by [27]. The current study used the original formula [6], because the number of bulls equal to a cow in terms of the reliability of GEBV must be compared under the same reliability of GEBV and is written without the term of the reliability of GEBV as shown in (9). The accuracies of the GEBV in this study are not comparable with the correlations between GEBV and degressed regression proofs determined from validation studies of field data [14,30]. The reason for this difference is that the information here is based on LD information alone, whereas the markers used for prediction of GEBV capture information on both genetic relationship and LD [24]. Both the theoretical number of bulls equal to a cow and the actual number derived from field data in terms of reliability warrants further study to validate the developed formula. However, the value of cows in terms of reliability of GEBV in the reference population under selection was simplified to the formula in (9), which likely will be a highly useful guideline for creating reference populations containing both bulls and cows.

CONCLUSION

Bias was greater for bulls than cows, because the intensity of preselection was higher in the bull population. Bias and reduction in reliability of GEBV due to preselection was alleviated by expanding reference population and by increasing the size of the reference population of cows even if the size of the reference population of bulls was held constant. Therefore, cows can contribute to reducing bias and increasing reliability due to their ease of use in expanding reference population size and by providing more recent animals compared with bulls. The number of bulls equal to a cow in a standpoint of bringing about the same size of reliabilities of the GEBV of preselected animals in the reference population was described as a simple formula (9) composed of reliability of the EBV of the trait of interest, preselection intensity and accuracy whether a reference population is either bulls/cows only or bulls and cows both. The generalized formulas presented in this study do satisfy the property of invariance and thus, is a general guideline for creating the reference population under selection for any combination of bull and cow populations.

CONFLICT OF INTEREST

We certify that there is no conflict of interest with any financial organization regarding the material discussed in the manuscript.

ACKNOWLEDGMENTS

We thank our teacher, Lin CY for helpful discussion and constructive comments and wish to thank the reviewers for the constructive comments which help improve the manuscript.

REFERENCES

1. Meuwissen THE, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. Genetics 2001;157:1819-29.
2. Mäntysaari E, Z Liu, vanRaden PM. Interbull validation test for genomic evaluations. Interbull bulletin 2010;41:4-5.
3. Falconer DS, Mackay TFC. Introduction to quantitative genetics. 4th ed. Harlow, UK: Longman; 1996.
4. Schaeffer LR. Strategy for applying genome-wide selection in dairy cattle. J Anim Breed Genet 2006;123:218-23.
5. Van Grevenhof EM, Van Arendonk JAM, Bijma P. Response to genomic selection: The bulmer effect and the potential of genomic selection when the number of phenotypic records is limiting. Genet Sel Evol 2012;44:26.
6. Daetwyler HD, Villanueva B, Bijma P, et al. Inbreeding in genome-wide selection. J Anim Breed Genet 2007;124:369-76.
7. Jenko J, Wiggans GR, Cooper TA, et al. Cow genotyping strategies for genomic selection in a small dairy cattle population. J Dairy Sci 2017;100:439-52.
8. Buch LH, Kargo M, Berg P, et al. The value of cows in reference

populations for genomic selection of new functional traits. Animal 2012;6:880-6.
9. Ding X, Zhang Z, Li X, et al. Accuracy of genomic prediction for milk production traits in the Chinese Holstein population using a reference population consisting of cows. J Dairy Sci 2013;96:5315-23.
10. Uemoto Y, Osawa T, Saburi J. Effect of genotyped cows in the reference population on the genomic evaluation of Holstein cattle. Animal 2017;11:382-93.
11. Calus MPL, de Haas Y, Veerkamp RF. Combining cow and bull reference populations to increase accuracy of genomic prediction and genome-wide association studies. J Dairy Sci 2013;96:6703-15.
12. Pryce JE, Nguyen TTT, Axford M, et al. Symposium review: Building a better cow—the Australian experience and future perspectives. J Dairy Sci 2018;101:3702-13.
13. Cunningham EP. Multi-stage index selection. Theor Appl Genet 1975;46:55-61.
14. Van Raden PM, Van Tassell C, Wiggans P, et al. Invited review: Reliability of genomic predictions for North American Holstein bulls. J Dairy Sci 2009;92:16-24.
15. Olson KM, van Raden PM, Tooker ME, et al. Differences among methods to validate genomic evaluations for dairy cattle. J Dairy Sci 2011;94:2613-20.
16. Daetwyler HD, Villanueva B, Woolliams JA. Accuracy of predicting the genetic risk of disease using a genome-wide approach. PLoS One 2008;3:e3395.
17. Hayes BJ, Daetwyler HD, Bowman PJ, et al. Accuracy of genomic selection: comparing theory and results. Proc Assoc Advmt Anim Breed Genet 2009;18:34-37.
18. Garrick DJ, Taylor JF, Fernando RL. Deregressing estimated breeding values and weighting information for genomic regression analyses. Genet Sel Evol 2009;41:55.
19. Liu Z, Seefried FR, Reinhardt F, et al. Impacts of both reference population size and inclusion of a residual polygenic effect on the accuracy of genomic prediction. Genet Sel Evol 2011;43:19.
20. Gonzalez-Recio O, Pryce JE, Haile-Mariam H, et al. Incorporating heifer feed efficiency in the Australian selection index using genomic selection. J Dairy Sci 2014;97:3883-93.
21. Weigel KA, Hoffman PC, Herring W. Potential gains in lifetime net merit from genomic testing of cows, heifers, and calves on commercial dairy farms. J Dairy Sci 2012;95:2215-25.
22. Thomasen JR, Sørensen AC, Lund MS, et al. Adding cows to the reference population makes a small dairy population competitive. J Dairy Sci 2014;97:5822-32.
23. Pryce JE, Nguyen TTT, Axford M, et al. Symposium review: Building a better cow—The Australian experience and future perspectives. J Dairy Sci 2018;101:3702-13.
24. Habier D, Fernando RL, Dekkers JCM. The impact of genetic relationship information on genome-assisted breeding values. Genetics 2007;177:2389-97.
25. Habier D, Tetens J, Seefried FR, et al. The impact of genetic relationship information on genomic breeding values in German Holstein cattle. Genet Sel Evol 2010;42:5.
26. Su G, Ma P, Nielsen US, et al. Sharing reference data and including cows in the reference population improve genomic predictions in Danish Jersey. Animal 2016;10:1067-75.
27. Goddard ME. Genomic selection: prediction of accuracy and maximisation of long term response. Genetica 2009;136:245-57.
28. Solberg TR, Sonesson AK, Woolliams JA, et al. Genomic selection using different marker types and densities. J Anim Sci 2008;86:2447-54.
29. Muir, WM. Comparison of genomic and traditional BLUP-estimated breeding value accuracy and selection response under alternative trait and genomic parameters. J Anim Breed Genet 2007;124:342-55.
30. Su G, Madsen P, Nielsen US, et al. Genomic prediction for Nordic red cattle using one-step and selection index blending. J Dairy Sci 2012;95:909-17.