J. Dairy Sci. 104
https://doi.org/10.3168/jds.2020-19789
© 2021, The Authors. Published by Elsevier Inc. and Fass Inc. on behalf of the American Dairy Science Association®.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Genomic predictions for yield traits in US Holsteins with unknown parent groups

A. Cesarani,[1]* Y. Masuda,[1] S. Tsuruta,[1] E. L. Nicolazzi,[2] P. M. VanRaden,[3] D. Lourenco,[1] and I. Misztal[1]
[1]Department of Animal and Dairy Science, University of Georgia, Athens 30602
[2]Council on Dairy Cattle Breeding, Bowie, MD 20716
[3]Animal Genomics and Improvement Laboratory, Agricultural Research Service, USDA, Beltsville, MD 20705-2350

ABSTRACT

The objective of this study was to assess the reliability and bias of estimated breeding values (EBV) from traditional BLUP with unknown parent groups (UPG), genomic EBV (GEBV) from single-step genomic BLUP (ssGBLUP) with UPG for the pedigree relationship matrix (**A**) only (SS_UPG), and GEBV from ssGBLUP with UPG for both **A** and the relationship matrix among genotyped animals ($\mathbf{A}_{22}$; SS_UPG2) using 6 large phenotype-pedigree truncated Holstein data sets. The complete data included 80 million records for milk, fat, and protein yields from 31 million cows born since 1980. Phenotype-pedigree truncation scenarios included truncation of phenotypes for cows born before 1990 and 2000 combined with truncation of pedigree information after 2 or 3 ancestral generations. A total of 861,525 genotyped bulls with progeny and cows with phenotypic records were used in the analyses. Reliability and bias (inflation/deflation) of GEBV were obtained for 2,710 bulls based on deregressed proofs, and on 381,779 cows born after 2014 based on predictivity (adjusted cow phenotypes). The BLUP reliabilities for young bulls varied from 0.29 to 0.30 across traits and were unaffected by data truncation and number of generations in the pedigree. Reliabilities ranged from 0.54 to 0.69 for SS_UPG and were slightly affected by phenotype-pedigree truncation. Reliabilities ranged from 0.69 to 0.73 for SS_UPG2 and were unaffected by phenotype-pedigree truncation. The regression coefficient of bull deregressed proofs on (G)EBV (i.e., GEBV and EBV) ranged from 0.86 to 0.90 for BLUP, from 0.77 to 0.94 for SS_UPG, and was 1.00 ± 0.03 for SS_UPG2. Cow predictivity ranged from 0.22 to 0.28 for BLUP, 0.48 to 0.51 for SS_UPG, and 0.51 to 0.54 for SS_UPG2. The highest cow predictivities for BLUP were obtained with the most extreme truncation, whereas for SS_UPG2, cow predictivities were also unaffected by phenotype-pedigree truncations. The regression coefficient of cow predictivities on (G)EBV was 1.02 ± 0.02 for SS_UPG2 with the most extreme truncation, which indicated the least biased predictions. Computations with the complete data set took 17 h with BLUP, 58 h with SS_UPG, and 23 h with SS_UPG2. The same computations with the most extreme phenotype-pedigree truncation took 7, 36, and 15 h, respectively. The SS_UPG2 converged in fewer rounds than BLUP, whereas SS_UPG took up to twice as many rounds. Thus, the ssGBLUP with UPG assigned to both **A** and $\mathbf{A}_{22}$ provided accurate and unbiased evaluations, regardless of phenotype-pedigree truncation scenario. Old phenotypes (before 2000 in this data set) did not affect the reliability of predictions for young selection candidates, especially in SS_UPG2.

Key words: unknown parent group, reliabilities, genomic selection, cow validation

INTRODUCTION

Genomic evaluations for the US dairy industry follow the multistep method presented in VanRaden (2008) and VanRaden et al. (2009). In this method, BLUP is used to derive deregressed proofs or equivalent pseudo-phenotypes, which are then used as dependent variables in a model to estimate SNP effects. Genomic PTA are obtained as weighted sums of SNP estimated values times the number of reference alleles in SNP genotypes and combined with parent averages adjusted for the portion due to genotyped animals to avoid double counting. This method is particularly appropriate when phenotypic and genomic data have separate ownership, and it is well accepted by the dairy industry.

The multistep method was developed when genotyping was performed almost exclusively on bulls, and genomic selection was yet to be widespread. When genotyping was extended to a large number of females, modifications were developed to reduce bias in genomic EBV (**GEBV**; Wiggans et al., 2011, 2012), and lately,

Received October 14, 2020.
Accepted December 18, 2020.
*Corresponding author: alberto.cesarani@uga.edu

the modifications have been replaced by a reduction in heritabilities (VanRaden et al., 2014). Additionally, with the use of genomic selection, deregressed proofs from BLUP became biased by preselection (Patry and Ducrocq, 2011). Adjustments for preselection are ongoing, following the amount of preselection by year and sex.

A valid alternative to the multistep procedure is single-step genomic BLUP that utilizes phenotypic, pedigree, and genomic information (Legarra et al., 2009; Christensen and Lund, 2010). This method is an extension of the traditional BLUP, where the pedigree relationship matrix (**A**) is replaced by **H**, the realized relationship matrix that blends **A** and the genomic relationship matrix **G** (Aguilar et al., 2010). Single-step GBLUP can use male and female genotypes directly and accounts for genomic preselection (Masuda et al., 2018). Although single-step genomic BLUP (**ssGBLUP**) was initially expensive with a large number of genotypes, computing refinements exploiting small effective population size and limited dimensionality of the genomic information (15,000 for Holstein) made computing inexpensive for any population size (Fragomeni et al., 2015; Masuda et al., 2016; Misztal, 2016).

Past applications of ssGBLUP in dairy were limited because of problems with unknown parent groups (**UPG**) used to account for missing pedigree information. When UPG were applied only to **A**, the convergence of ssGBLUP was slow, or the iterative method for obtaining solutions diverged (Tsuruta et al., 2014; Matilainen et al., 2016). Misztal et al. (2013) proposed using UPG for all components of the inverse of **H** with the Quaas–Pollak (**QP**) transformation (Quaas, 1988). The new UPG strategy improved the convergence rate and was successfully applied by Matilainen et al. (2016) in a Nordic dairy cattle population, and by Misztal et al. (2017) in US Holsteins with data recorded until 2014. Masuda et al. (2018) analyzed the US Holstein data set with records until 2015 and found lower reliabilities for GEBV from ssGBLUP with the new UPG compared with a model with no UPG. These UPG were poorly estimated, likely because a large number of genotyped commercial females born after 2014 had missing or incorrect pedigree information, and most of them had no phenotypes (Bradford et al., 2019). Tsuruta et al. (2019) found that UPG could be better estimated by refining the UPG formulas in ssGBLUP, reducing the number of groups, and incorporating inbreeding for unknown parents.

A comprehensive approach for the construction of UPG was presented by Legarra et al. (2015), using the concept of metafounders, in which incomplete pedigree relationships are adjusted to **G** derived with equal gene frequencies using various algorithms. In subsequent studies, ssGBLUP with metafounders delivered the least biased GEBV in simulated data sets, whereas ssGBLUP with UPG as in Tsuruta et al. (2019) was less biased in field data sets. Most likely, the use of metafounders requires updated parameter estimation to work well with highly unbalanced data sets.

Truncating old pedigree information can also help to reduce biases due to missing information (Lourenco et al., 2014). In fact, an important question in genetic evaluations is whether old generations of pedigree, phenotypes, and genotypes should be used. Despite access to a large number of generations, some broiler companies use only 3 generations of data because more generations do not improve prediction accuracy for selection candidates but do increase computing time (Vivian Breen, Cobb-Vantress Inc., Siloam Springs, AR, personal communication). Lourenco et al. (2014) looked into the effect of data and pedigree truncation on accuracy of genomic predictions in dairy cattle and pigs. Truncation to about 2 generations of data did not reduce accuracy and sometimes improved it. Similar results were found by Howard et al. (2018). Data truncation in dairy cattle may be affected by decreasing generation intervals (Mäntysaari et al., 2020), change of trait definitions under improved management (Tsuruta et al., 2005), and decay of genomic information with selection (Muir, 2007).

The objective of this study was to assess the reliability and bias of EBV from traditional BLUP with UPG, GEBV from ssGBLUP with UPG for the pedigree relationship matrix (**A**) only (**SS_UPG**), and GEBV from ssGBLUP with UPG for both **A** and the relationship matrix among genotyped animals ($\mathbf{A}_{22}$; **SS_UPG2**) using 6 large phenotype-pedigree truncated Holstein data sets.

MATERIALS AND METHODS

Data

Data used in the official genetic evaluations for yield traits in US Holstein cattle were provided by the Council on Dairy Cattle Breeding (Bowie, MD). Three different phenotype cut-off dates were considered: (1) yield records recorded from January 1980 to December 2018 (**Pheno1980**); (2) records from January 1990 to December 2018 (**Pheno1990**); and (3) records from January 2000 to December 2018 (**Pheno2000**). In addition, 2 pedigree depths that traced pedigree information for 2 (**Depth2**) and 3 generations (**Depth3**) back for animals with phenotypes, genotypes, or both, were considered for each phenotypic data set. This yielded 6 phenotype-pedigree truncation scenarios: Pheno1980-Depth3, Pheno1980-Depth2, Pheno1990-

Table 1. Number of records, cows, genotypes, and pedigree animals in 3 phenotype cut-off and 2 pedigree depth scenarios

Phenotype cut-off scenario[1]	Phenotype cut-off year	Number of		Genotypes[2]	Animals in pedigree[3]	
		Records	Cows		Depth3	Depth2
Pheno1980	1980	77,799,792	31,485,126	861,525	40,527,946	39,879,498
Pheno1990	1990	61,229,782	25,694,479	861,525	34,019,303	33,252,617
Pheno2000	2000	42,218,748	18,373,166	861,525	25,173,433	24,406,750

[1]Pheno1980 = phenotypes from animals born between January 1980 and December 2018; Pheno1990 = phenotypes from January 1990 to December 2018; and Pheno2000 = phenotypes from January 2000 to December 2018.

[2]Genotypes were used in genomic models only; that is, SS_UPG (SsGBLUP with unknown parent groups only in **A**) and SS_UPG2 (ssGBLUP with unknown parent groups for both **A** and the pedigree relationship matrix for genotyped animals).

[3]Depth3 = pedigree traced 3 generations back; Depth2 = pedigree traced 2 generations back.

Depth3, Pheno1990-Depth2, Pheno2000-Depth3, and Pheno1990-Depth2. Genotypes for 79,294 SNP from 861,525 animals were included in the analyses of the 6 phenotype-pedigree truncation scenarios (613,210 cows with phenotypes plus 248,315 ancestors of these cows). Number of records, cows, and genotypes in the 3 phenotype cut-offs (1980, 1990, and 2000) as well as number of pedigree animals in the 6 phenotype-pedigree truncation scenarios are shown in Table 1. Numbers of cows with both phenotypes and genotypes changed among scenarios; there were 613,210 in Pheno1980, 613,154 in Pheno1990, and 612,148 in Pheno2000, respectively.

Model and Analysis

Three evaluation methods were considered: (1) traditional BLUP with unknown parent groups based on year of birth, (2) ssGBLUP with unknown parent groups only in **A** (SS_UPG), and (3) ssGBLUP with UPG for both **A** and the pedigree relationship matrix for genotyped animals ($\mathbf{A}_{22}$) (SS_UPG2), implemented as in Tsuruta et al. (2019). In all the 3 models, UPG were defined according to year of birth, which ranged from 1974 to 2018. As one UPG was set every 2 yr, a total of 23 UPG were used in the analyses. The number of animals with missing parents that were linked to each UPG is in Supplementary File S1 (http://dx.doi.org/10.17632/r3v49629jf.1; Cesarani, 2021); the number of animals with phenotypes and linked to UPG is in Supplementary File S2.

The 3-trait repeatability animal model used for the traditional BLUP model was

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Z_h h} + \mathbf{Z_a Q_a g_a} + \mathbf{Z_a a} + \mathbf{Z_p p} + \mathbf{e},$$

where $\mathbf{y}$ = vector of milk, fat, and protein yield phenotypic records; $\mathbf{b}$ = vector of fixed herd-management effects, age-parity effects, inbreeding coefficient covariates, and Holstein fraction covariates; $\mathbf{h}$ = vector of random herd-sire effects; $\mathbf{g_a}$ = vector of unknown parent groups based on year of birth; $\mathbf{a}$ = vector of animal additive genetic effects; $\mathbf{p}$ = vector of permanent environmental effects; and $\mathbf{e}$ = vector of residuals. Matrix $\mathbf{X}$ is an incidence matrix relating phenotypic records in vector $\mathbf{y}$ to fixed effects in vector $\mathbf{b}$, matrix $\mathbf{Z_h}$ is an incidence matrix relating phenotypic records in vector $\mathbf{y}$ to herd-sire effects in vector $\mathbf{h}$, matrix $\mathbf{Z_a}$ is an incidence matrix relating phenotypic records in vector $\mathbf{y}$ to animal additive genetic effects in vector $\mathbf{a}$, matrix $\mathbf{Q_a}$ is an incidence matrix relating animals in vector $\mathbf{a}$ to unknown parent groups in vector $\mathbf{g_a}$, and matrix $\mathbf{Z_p}$ is an incidence matrix relating phenotypic records in vector $\mathbf{y}$ to permanent environmental effects in vector $\mathbf{p}$. Vector $\mathbf{h}$ has mean zero and variance $\mathbf{I} \otimes \mathbf{V_h}$, where $\mathbf{V_h}$ is a diagonal 3×3 matrix of herd-sire variances and covariances, and $\otimes$ is the Kronecker product. Vector $\mathbf{a}$ has mean zero and variance $\mathbf{\Lambda} \otimes \mathbf{V_a}$, where $\mathbf{\Lambda}$ is $\mathbf{A}$ in BLUP and $\mathbf{H}$ in ssGBLUP and $\mathbf{V_a}$ is a 3×3 matrix of additive genetic variances and covariances. Vector $\mathbf{p}$ has mean zero and variance $\mathbf{I} \otimes \mathbf{V_p}$, where $\mathbf{V_p}$ is a 3×3 matrix of permanent environment variances and covariances. Vector $\mathbf{e}$ has mean zero and variance $\mathbf{I} \otimes \mathbf{V_e}$, where $\mathbf{V_e}$ is a 3×3 matrix of residual variances and covariances.

The 3-trait SS_UPG and SS_UPG2 models contained the same effects as the traditional BLUP model; however, the covariance structure for $\mathbf{Z_a a}$ and $\mathbf{Z_a Q_a g_a}$ included genomic relationships. The covariance matrix for the additive genetic effect in SS_UPG was given by

$$\mathbf{H}^*_{UPG} = \mathbf{A}^* + \begin{bmatrix} 0 & 0 & 0 \\ 0 & \mathbf{G}^{-1}_{\mathrm{APY}} - \mathbf{A}^{-1}_{22} & -\left(\mathbf{G}^{-1}_{\mathrm{APY}} - \mathbf{A}^{-1}_{22}\right)\mathbf{Q}_2 \\ 0 & -\mathbf{Q}_2{}'\left(\mathbf{G}^{-1}_{\mathrm{APY}} - \mathbf{A}^{-1}_{22}\right) & \mathbf{Q}_2{}'\left(\mathbf{G}^{-1}_{\mathrm{APY}} - \mathbf{A}^{-1}_{22}\right)\mathbf{Q}_2 \end{bmatrix},$$

where $\mathbf{H}^*_{UPG}$ is the inverse of the realized relationship matrix with unknown parent groups added to the pedigree relationship matrix $\mathbf{A}$, to the matrix of pedigree relationships among genotyped animals $\mathbf{A}_{22}$, and to the

genomic relationship matrix $\mathbf{G}$, which was built according to method 1 of VanRaden (2008), being centered and scaled by current allele frequencies from all genotyped animals; $\mathbf{A}^*$ is the inverse of the pedigree relationship matrix with unknown parent groups, i.e., modified with the QP transformation (Quaas, 1988); $\mathbf{G}_{\text{APY}}^{-1}$ is the inverse of the genomic relationship matrix constructed using the algorithm for proven and young (APY; Misztal et al., 2014) with 15,000 randomly selected core animals; $\mathbf{A}_{22}^{-1}$ is the inverse of the matrix of pedigree relationships among genotyped animals, and $\mathbf{Q}_2$ is a matrix assigning genotyped animals to UPG. The covariance matrix for the additive genetic effect in SS_UPG2 was given by

$$\mathbf{H}_{UPG2}^{*} = \mathbf{A}^{*} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & \mathbf{G}_{\text{APY}}^{-1} - \mathbf{A}_{22}^{-1} & -\mathbf{A}_{22}^{-1}\mathbf{Q}_2 \\ 0 & -\mathbf{Q}_2{}'\mathbf{A}_{22}^{-1} & \mathbf{Q}_2{}'\mathbf{A}_{22}^{-1}\mathbf{Q}_2 \end{bmatrix},$$

where $\mathbf{H}_{UPG2}^{*}$ is the inverse of the realized relationship matrix with unknown parent groups added to the pedigree relationship matrix $\mathbf{A}$, to the matrix of pedigree relationships among genotyped animals $\mathbf{A}_{22}$; all the matrices have been previously described. As $\mathbf{G}_{\text{APY}}^{-1}$ relies on the inverse of $\mathbf{G}$ for core animals, 95% of $\mathbf{G}$ was combined to 5% of $\mathbf{A}_{22}$ to avoid singularity problems.

The 3-trait mixed model equations were solved using iteration on data and a block preconditioning conjugate gradient with software the BLUP90IOD2 (version 3.113; Tsuruta et al., 2001; Tsuruta and Misztal, 2008). Variance components for milk yield, fat yield, and protein yield were obtained from VanRaden et al. (2014). The heterogeneous herd variance was adjusted according to Wiggans and VanRaden (1991). Computations were carried out on a Linux server (x86_64) with 512 GB of RAM and equipped with Intel Xeon E5–2683 v4 (2.10 GHz) processors with 32 computing cores. The convergence criterion was 10^{-15}, and was computed as the norm squared of the difference between the right-hand-side and the left-hand-side of the mixed model equations, scaled by the norm squared of left-hand-side.

Validation of Breeding Values

Genomic EBV were validated for each phenotype-pedigree truncation scenario. Phenotypes for cows born after 2014 were removed in the reduced data sets. Genotyped cows born between 2014 and 2017 that had phenotypes in the complete but not in the reduced data sets were considered to be validation cows, whereas validation bulls were the genotyped sires of those validation cows (i.e., validation bulls without daughters with phenotypes in the reduced data sets).

The ability to predict the future performance of young candidates was used for validation (Tsuruta et al., 2011; Lourenco et al., 2015; Masuda et al., 2016). Estimates of daughter yield deviations (**DYD**) for validation bulls and phenotypes adjusted for all fixed effects and random effects other than animal and residual ($\mathbf{Y_{PRED}}$) for validation cows were obtained using the complete data set. The DYD were estimated for all bulls using the method by Liu et al. (2004) and the algorithm by Mrode and Swanson (2004), both implemented using in-house software. The Y_{PRED} were calculated using PREDICTF90 (version 1.5; Misztal et al., 2018).

Regression coefficients of DYD or Y_{PRED} on (G)EBV (i.e., GEBV and EBV) were used to measure the inflation of predictions for 2,710 bulls and 381,779 cows in the validation data set. The coefficient of determination between DYD and (G)EBV was used as a measure of reliability of predictions for validation bulls. The reliability for validation cows was calculated as the square of the correlation between Y_{PRED} and (G)EBV divided by the heritability; the simple correlation is known as predictive ability.

Last, the stability of subsequent BLUP, SS_UPG, and SS_UPG2 evaluations for validation bulls and cows was calculated as the correlation between predictions from the reduced and complete data sets (Legarra and Reverter, 2018).

RESULTS AND DISCUSSION

Bull Reliabilities

Reliabilities of (G)EBV for bulls with the 3 methods and the 6 phenotype-pedigree truncation scenarios are shown in Figure 1. Bull EBV reliabilities from BLUP were low (around 0.30) and similar across the 6 phenotype-pedigree truncation scenarios. The small effect of past generations could be explained by the fact that the BLUP EBV for validation animals were equal to parent averages, and high-reliability sires and dams were chosen as potential sires of bulls. Bull GEBV reliabilities from SS_UPG were as high as 0.69, and interaction between phenotype cut-off year and pedigree depth was observed. Bull GEBV reliabilities from SS_UPG2 declined with fewer phenotypes. Conversely, the reliabilities with pedigree depth3 were slightly higher with less data. Bull reliabilities from SS_UPG2 were highest (up to 0.73), and changes due to year of phenotype cut-off were less numerous. The lack of SS_UPG2 reliability interactions between phenotype cut-offs and pedigree

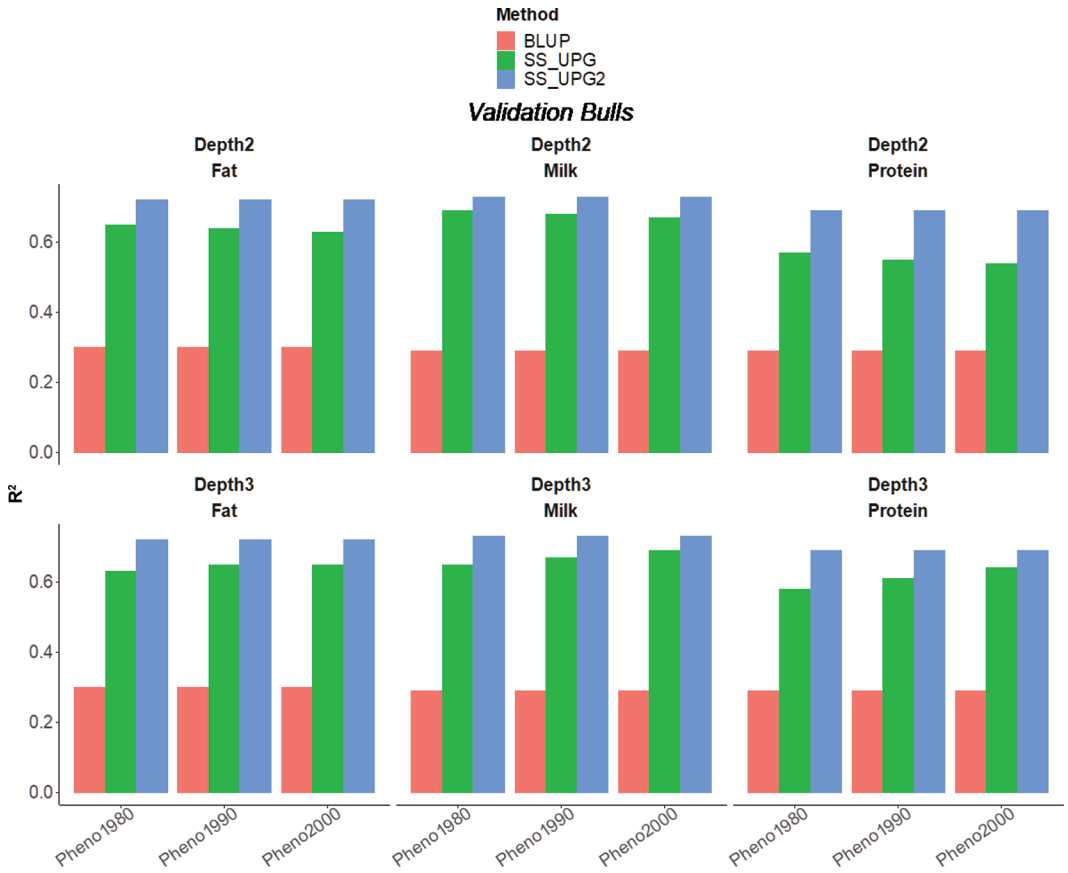

Figure 1. Coefficients of determination (R^2) between bull daughter yield deviations (DYD) and EBV from BLUP in 2019 and between DYD and genomic EBV from SS_UPG (ssGBLUP with unknown parent groups only in **A**) and SS_UPG2 (ssGBLUP with unknown parent groups for both **A** and the pedigree relationship matrix for genotyped animals) in 2014 in the validation data set under 6 phenotype-pedigree truncation scenarios. Pheno1980 = phenotypes from animals born between January 1980 and December 2018; Pheno1990 = phenotypes from January 1990 to December 2018; and Pheno2000 = phenotypes from January 2000 to December 2018. Depth3 = pedigree traced 3 generations back; Depth2 = pedigree traced 2 generations back.

depths indicated that this method was effective in accounting for a base population when data were missing.

Slightly higher coefficients of determination were observed for milk yield, followed by fat and protein yields. Coefficient of determination differences among traits were larger with SS_UPG and SS_UPG2 than with BLUP. The highest bull coefficients of determination from SS_UPG2 ranged from 0.69 (protein yield) to 0.73 (milk yield). These raw reliability values were similar to the adjusted reliabilities of 0.74 for milk yield, 0.73 for fat yield, and 0.66 for protein yield reported by VanRaden et al. (2017). Fragomeni et al. (2019) showed that the adjustment caused an average increase of 20% in reliabilities.

Masuda et al. (2018) applied the exact UPG (UPG assigned to **A**, **G**, and $\mathbf{A}_{22}$; Misztal et al., 2013) to the 2015 US Holstein data set with about 764,000 genotyped animals. They obtained reliabilities for protein yield with exact UPG that were 60% lower than with no UPG, possibly because the QP-transformed **H** with exact UPG was ill-conditioned. As **G** does not rely on pedigree missingness, there may be no need to fit UPG in **G** (Tsuruta et al., 2019). The latter authors reported less bias in genetic trends for type traits in Holsteins when UPG were fit only for **A** and $\mathbf{A}_{22}$; however, reliabilities were similar.

Cow Predictive Abilities and Reliabilities

Cow predictive ability (predictivity) and reliability from all methods and phenotype-pedigree truncation scenarios are presented in Table 2. The predictivities for BLUP ranged from 0.22 to 0.28 and were independent of pedigree depth, but they increased with year of phenotype cut-off. The reasons for the increase in predictivity with less data are unclear but may be due to preferential treatment of genotyped cows and inflated evaluation of ancestors. Cow predictivities with

Table 2. Predictive abilities and reliabilities (in parentheses) from BLUP, SS_UPG, and SS_UPG2 for validation cows (n = 381,779) under 6 phenotype-pedigree truncation scenarios[1]

	Validation cows								
	BLUP			SS_UPG			SS_UPG2		
Truncation scenario[2]	Milk	Fat	Protein	Milk	Fat	Protein	Milk	Fat	Protein
Pheno1980–Depth3	0.23	0.25	0.22	0.51	0.49	0.48	0.54	0.54	0.51
	(0.14)	(0.14)	(0.13)	(0.74)	(0.69)	(0.66)	(0.83)	(0.83)	(0.74)
Pheno1980–Depth2	0.23	0.25	0.22	0.52	0.50	0.48	0.54	0.54	0.51
	(0.14)	(0.14)	(0.13)	(0.77)	(0.71)	(0.66)	(0.83)	(0.83)	(0.74)
Pheno1990–Depth3	0.24	0.25	0.23	0.51	0.49	0.48	0.54	0.54	0.51
	(0.14)	(0.14)	(0.14)	(0.74)	(0.69)	(0.66)	(0.83)	(0.83)	(0.74)
Pheno1990–Depth2	0.24	0.26	0.23	0.51	0.49	0.48	0.54	0.54	0.51
	(0.14)	(0.15)	(0.14)	(0.74)	(0.69)	(0.66)	(0.83)	(0.83)	(0.74)
Pheno2000–Depth3	0.26	0.28	0.25	0.51	0.49	0.48	0.54	0.54	0.51
	(0.15)	(0.15)	(0.14)	(0.74)	(0.69)	(0.66)	(0.83)	(0.83)	(0.74)
Pheno2000–Depth2	0.26	0.28	0.25	0.50	0.48	0.49	0.54	0.54	0.51
	(0.15)	(0.15)	(0.14)	(0.71)	(0.66)	(0.69)	(0.83)	(0.83)	(0.74)

[1]Heritability was assumed to be 0.35 for the computation of reliabilities. SS_UPG = ssGBLUP with unknown parent groups only in **A**; SS_UPG2 = ssGBLUP with unknown parent groups for both **A** and the pedigree relationship matrix for genotyped animals.

[2]Pheno1980 = phenotypes from animals born between January 1980 and December 2018; Pheno1990 = phenotypes from January 1990 to December 2018; and Pheno2000 = phenotypes from January 2000 to December 2018. Depth3 = pedigree traced 3 generations back; Depth2 = pedigree traced 2 generations back.

SS_UPG reached 0.51 and were stable with pedigree depth3. Cow predictivities with pedigree depth3 were slightly higher, with more data for milk and fat yields but slightly less for protein yield. Cow predictivities with SS_UPG2 were similar across all truncation scenarios. Variation in SS_UPG cow predictivity among different phenotype-pedigree truncation scenarios were likely due to the inability of SS_UPG to account for UPG over time.

Average cow predictivities across traits were 0.25 ± 0.02 for BLUP, 0.49 ± 0.01 for SS_UPG, and 0.53 ± 0.01 for SS_UPG2, whereas average cow reliabilities were 0.17 ± 0.02 for BLUP, 0.70 ± 0.04 for SS_UPG, and 0.80 ± 0.04 for SS_UPG2. The increases of 8% in cow predictivity and 14% in cow reliability achieved by fitting UPG to both **A** and $\mathbf{A}_{22}$ instead of only to **A** show the importance of correctly accounting for missing pedigree information in ssGBLUP. Thus, SS_UPG2 was the most appropriate method to describe prediction changes over time and to estimate future yield performance of cows. Almost no differences in cow predictivity were observed among the 6 phenotype-pedigree truncation scenarios with SS_UPG2; however, differences existed among traits. On average, the lowest cow predictivities were those for protein yield and the largest were for milk yield with SS_UPG and SS_UPG2 and for fat yield with BLUP.

Predictivity can be tied to reliability. If $\text{accuracy} = \text{predictivity}/\sqrt{h^2}$ (Legarra et al., 2008), then $\text{reliability} = \text{predictivity}^2/h^2$, where h^2 is the heritability of the trait. Cow predictivity reached 0.54 in this study. This means that for reliability < 1, $h^2 > \text{predictivity}^2$ or $h^2 > 0.54^2$; hence, h^2 must be greater than 0.29. However, this argument assumes that only one record per cow is available. In our study, validation cows had an average of 1.46 ± 0.63 lactation records in the complete data. Some studies have indicated that heritabilities for dairy traits are declining, with low heritabilities providing the highest reliabilities and lowest biases for GEBV (VanRaden et al., 2014). Tsuruta et al. (2019) reported that GEBV inflation can be reduced by using smaller additive genetic variances in the analysis. However, this may be an artifact due to using BLUP in the first stage of the evaluation, or perhaps because of inaccurate measures of reliability or bias. The algorithm used here to compute DYD for sires depends on solutions obtained by BLUP; hence, it may be biased because it does not account for selection decisions based on genomic information (Patry and Ducrocq, 2011; Masuda et al., 2018). Further, cow predictivity is a function of solutions for fixed effects, which may be biased because of incomplete model fitting (Bermann et al., 2020).

Biases for Bull and Cow (G)EBV and Consistency Between Subsequent Evaluations

Regression coefficients of validation bull DYD on (G)EBV for the 6 phenotype-pedigree truncation scenarios are shown in Figure 2. An interaction between phenotype cut-off year and pedigree depth is visible. The regression coefficients for validation bulls from BLUP and SS_UPG were very similar for all traits across the 6 phenotype-pedigree truncation scenarios, with values

close to 0.9, except for protein yield, which was around 0.85, indicating some degree of inflation. Masuda et al. (2018) reported a considerable level of inflation on GEBV for protein yield in US Holstein bulls when using ssGBLUP with UPG either in $\mathbf{A}$, $\mathbf{G}$, and $\mathbf{A}_{22}$ (regression coefficient = 0.51) or in $\mathbf{A}$ only (regression coefficient = 0.78). Conversely, the SS_UPG2 regression coefficients for validation bulls in our study, which considered UPG for $\mathbf{A}$ and $\mathbf{A}_{22}$, were very stable at 1.02 ± 0.02. Thus, correct fitting of UPG to account for missing pedigrees in ssGBLUP leads to less biased predictions.

The regression coefficients of cow Y_{PRED} on EBV were very low (around 0.5), indicating strong inflation. These low regression coefficients indicate that BLUP was inadequate to provide reasonable cow evaluations without any modification or rescaling, perhaps because of genomic preselection. The regression coefficient of cow Y_{PRED} on GEBV from SS_UPG was substantially higher (around 0.9), indicating that inflation was lower than with BLUP. The least inflated/deflated method was SS_UPG2 with a regression coefficient equal to 1.03 ± 0.03. Small fluctuations among regression coefficients were observed across the 6 phenotype-pedigree truncation scenarios.

Regression coefficients should be as close to 1 as possible, which means that (G)EBV for young validation animals are correctly shrunk, and therefore are good estimators of future male (DYD) and female (Y_{PRED}) performance. Values <1 indicate inflation, whereas higher values indicate deflation of (G)EBV. Deviations within 5% from unity are considered good predictions, and values within 15% are assumed to be acceptable (Tsuruta et al., 2011). Better BLUP regression coefficients (close to 1) were observed for bulls (0.89 ± 0.02) than for cows (0.50 ± 0.08). This was expected because of the larger amount of information available for males. Adding genomic information to the models improved the regression coefficients more for cows than for bulls. Overall, SS_UPG regression coefficients averaged 0.87 ± 0.06 for bulls and 0.93 ± 0.04 for cows. Better regression coefficients with ssGBLUP than with BLUP were reported for bulls from the 2014 US Holstein data set (around 569,000 genotyped animals; Masuda et al., 2016). The use of UPG in both $\mathbf{A}$ and $\mathbf{A}_{22}$ matrices led to better SS_UPG2 average regression coefficients for bulls (1.00 ± 0.02) and to slightly deflated breeding values for cows (1.06 ± 0.04). No large effects due to the 6 phenotype-pedigree truncation scenarios were found for validation bulls. However, a slight increasing trend was observed for validation cows because a reduction in the number of records had a positive effect on the validation of predictions. Average regression coefficient values for genomic and nongenomic methods were 0.80 ± 0.28 for Pheno1980, 0.82 ± 0.25 for Pheno1990, and 0.86 ± 0.21 for Pheno2000. The 2 considered pedigree depths (depth3 and depth2) had no effect on regression coefficients. Similar average regression coefficients were found for validation bulls and cows across all phenotype-pedigree truncation scenarios and prediction methods. This could be explained by small differences among the sizes of data sets in depth3 and depth2 (Table 1). On average, tracking the pedigree back for one more generation resulted in the addition of approximately 700,000 animals, representing only 2% of the population.

Correlations between (G)EBV estimated with the complete and reduced data sets are presented in Table 3. Higher values indicate that the considered method is more stable over time. The highest correlations were those from SS_UPG2, suggesting that (G)EBV based on reduced data sets using this method were good predictors of the (G)EBV of young animals when phenotypes of their daughters were added to the data set. Cows had higher correlations than bulls, possibly because they had less data added from one evaluation to the next. Correlations for BLUP and SS_UPG2 were invariant to phenotype cut-off year and pedigree depth, whereas small differences were observed for SS_UPG.

Computational Aspects

Table 4 shows the number of rounds to convergence and time per round for the 6 phenotype-pedigree truncation scenarios. Numbers of rounds to convergence were approximately 500 for BLUP, 1,000 for SS_UPG, and 400 for SS_UPG2. The smaller number of rounds of iteration with SS_UPG2 than with SS_UPG and BLUP indicates a better conditioned system of equations. As expected, the time per round was about twice as large for single-step methods than for BLUP. The total computing time was almost halved when only phenotypes from animals born from 2000 to 2018 (Pheno2000) were used, compared with using phenotypes from 1980 to 2018 (Pheno1980). Computation times to convergence with Pheno1980 took 17 h for BLUP, 58 h for SS_UPG, and 23 h for SS_UPG2. These computation times were reduced to 7 h for BLUP, 36 h for SS_UPG, and 15 h for SS_UPG2 with Pheno2000.

The lower computing times with SS_UPG2 than with SS_UPG indicate that the inverse of $\mathbf{H}$ was better conditioned when UPG were fitted to both $\mathbf{A}$ and $\mathbf{A}_{22}$ than to $\mathbf{A}$ only. Further, SS_UPG2 provided the greatest reliability and the least inflated/deflated GEBV for validation bulls and cows, and it was less affected by the removal of old data. The small influence of old phenotypes on reliability could be due to a massive influx of genomic information that made phenotypic data tied

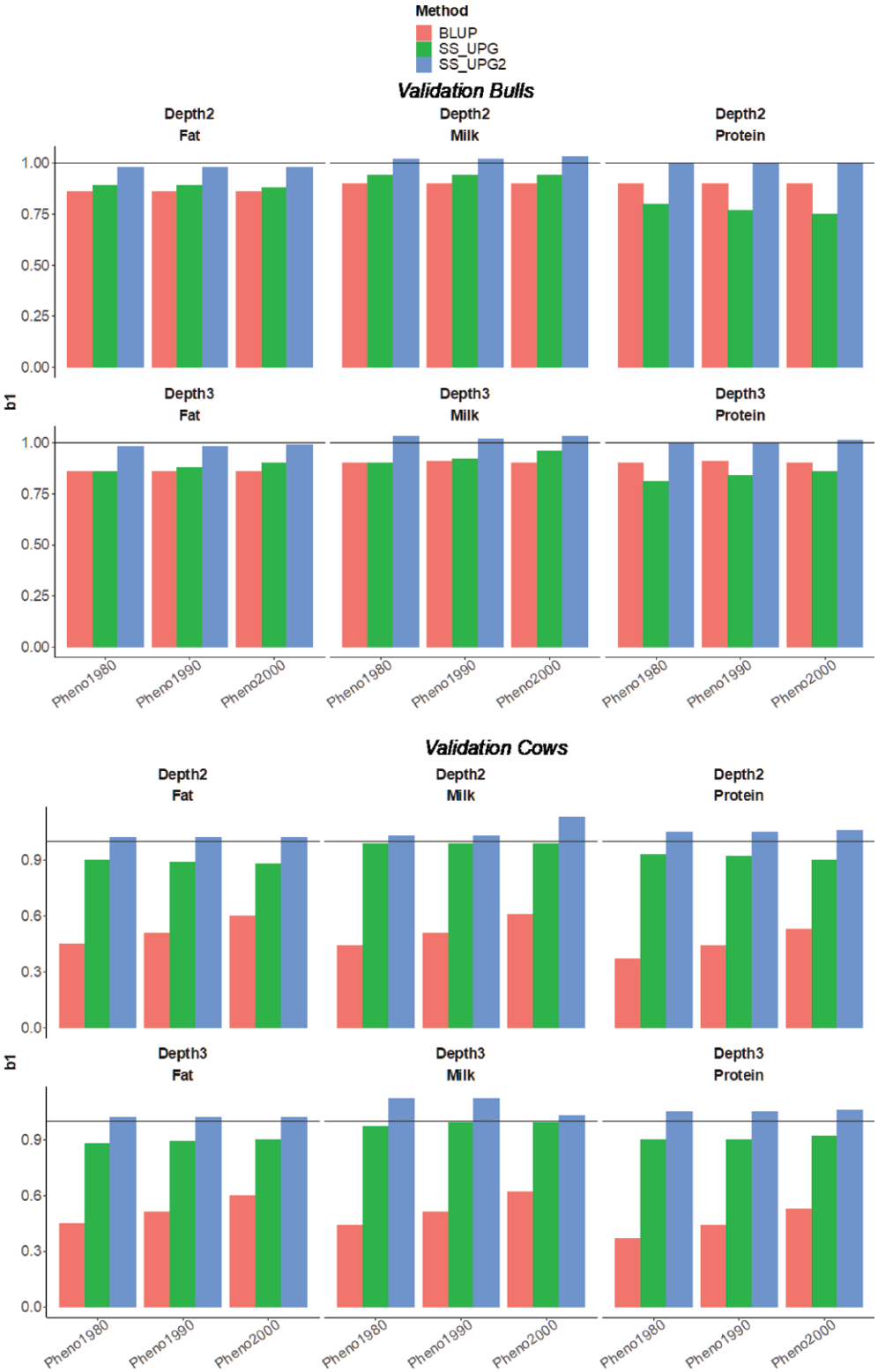

Figure 2. Regression coefficients (b1) of bull daughter yield deviations (DYD, top) and cow adjusted phenotypes (Y_{PRED}, bottom) on EBV from BLUP in 2018 and on genomic EBV from SS_UPG (ssGBLUP with unknown parent groups only in **A**) and SS_UPG2 (ssGBLUP with unknown parent groups for both **A** and the pedigree relationship matrix for genotyped animals) in 2014 in the validation data set under 6 phenotype-pedigree truncation scenarios. Pheno1980 = phenotypes from animals born between January 1980 and December 2018; Pheno1990 = phenotypes from January 1990 to December 2018; and Pheno2000 = phenotypes from January 2000 to December 2018. Depth3 = pedigree traced 3 generations back; Depth2 = pedigree traced 2 generations back.

Table 3. Correlations between bull (n = 2,710) and cow (n = 381,779) genomic EBV for milk, fat, and protein (Prot) estimated with BLUP, SS_UPG, and SS_UPG2 under 6 phenotype-pedigree truncation scenarios[1]

	BLUP						SS_UPG						SS_UPG2					
	Bulls			Cows			Bulls			Cows			Bulls			Cows		
Truncation scenario[2]	Milk	Fat	Prot	Milk	Fat	Prot	Milk	Fat	Prot	Milk	Fat	Prot	Milk	Fat	Prot	Milk	Fat	Prot
Pheno1980, Depth3	0.59	0.60	0.59	0.53	0.56	0.50	0.84	0.81	0.80	0.87	0.82	0.85	0.92	0.91	0.90	0.96	0.96	0.95
Pheno1980, Depth2	0.59	0.60	0.59	0.53	0.56	0.50	0.87	0.82	0.78	0.88	0.81	0.85	0.92	0.91	0.90	0.96	0.96	0.95
Pheno1990, Depth3	0.59	0.60	0.59	0.56	0.59	0.54	0.86	0.83	0.82	0.88	0.82	0.85	0.92	0.91	0.90	0.96	0.96	0.95
Pheno1990, Depth2	0.59	0.60	0.59	0.56	0.59	0.54	0.86	0.81	0.76	0.87	0.81	0.84	0.92	0.91	0.90	0.96	0.96	0.95
Pheno2000, Depth3	0.59	0.60	0.59	0.60	0.63	0.58	0.88	0.83	0.86	0.88	0.82	0.84	0.92	0.91	0.90	0.96	0.96	0.95
Pheno2000, Depth2	0.59	0.60	0.59	0.60	0.63	0.58	0.86	0.80	0.76	0.87	0.81	0.83	0.92	0.91	0.90	0.96	0.96	0.95

[1]SS_UPG = ssGBLUP with unknown parent groups only in **A**; SS_UPG2 = ssGBLUP with unknown parent groups for both **A** and the pedigree relationship matrix for genotyped animals.

[2]Pheno1980 = phenotypes from animals born between January 1980 and December 2018; Pheno1990 = phenotypes from January 1990 to December 2018; and Pheno2000 = phenotypes from January 2000 to December 2018. Depth3 = pedigree traced 3 generations back; Depth2 = pedigree traced 2 generations back.

to genomic information important, and pre-genomic phenotypes less important. The small effect of additional generations of pedigree could be linked to UPG correctly accounting for missing pedigree information for all animals including genotyped ones.

The current genomic evaluation for dairy cattle in the United States has 2 additional features not included in this study. First, it incorporates Interbull-provided deregressed proofs from other countries as external information. External information provided by Interbull was successfully included in ssGBLUP in 2019 (Guarini et al., 2019). Second, it uses a BayesA-type of approach (VanRaden, 2008) that considers different weights for SNP (including causative SNP), aiming to increase the reliability of predictions. Fragomeni et al. (2019) showed that causative SNP can be accommodated in ssGBLUP via a weighted genomic relationship matrix. In an evaluation of Belgian Blue cattle for type traits, Gualdrón-Duarte et al. (2020) showed that SNP selection/weighting in Bayesian methods or ssGBLUP increased the reliability on average from 0.46 to 0.50 compared with unweighted ssGBLUP. The reliability of unweighted ssGBLUP was 0.52. Utilization of weights from previous methods increased the reliability of weighted ssGBLUP to 0.55.

In our study, genomic models outperformed BLUP on all considered criteria (coefficient of determination, regression coefficient, consistency between subsequent evaluations), and between the genomic models, SS_UPG2 showed better performance than SS_UPG. Negligible differences were observed among the 3 phenotype cut-offs (Pheno1980, Pheno1990, and Pheno2000), indicating that decreasing the number of phenotypic records in the data set from ~78 to ~42 million did not affect the predictive ability of the models. The superior performance of SS_UPG2 indicated that missing pedigree information should be correctly modeled in both **A** and $\mathbf{A}_{22}$. Legarra et al. (2015) proposed the concept of metafounders, where UPG are treated as founder animals in the pedigree. This concept can be useful in situations in which pedigrees are complex or scarcely recorded, especially for multibreed populations. An ssGBLUP with metafounders improved the reliability and predictive ability of purebred and crossbred animals, and had less inflated/deflated predictions in a population of Landrace, Yorkshire, and Landrace-Yorkshire pigs using a joint pedigree with 2 metafounders to account for parental the breeds (Xiang et al., 2017).

CONCLUSIONS

Genomic predictions for yield traits in US Holstein bulls and cows using single-step GBLUP are reliable

Table 4. Number of rounds to convergence and computing times per round of iteration to obtain predicted values with BLUP, SS_UPG, and SS_UPG2 for 6 phenotype-pedigree truncation scenarios[1]

	Number of rounds			Computing time (s/round)		
Truncation scenario[2]	BLUP	SS_UPG	SS_UPG2	BLUP	SS_UPG	SS_UPG2
Pheno1980–Depth3	516	989	369	125	214	224
Pheno1980–Depth2	502	1,039	378	99	210	221
Pheno1990–Depth3	470	910	391	111	197	212
Pheno1990–Depth2	447	958	461	83	163	184
Pheno2000–Depth3	596	734	404	66	147	160
Pheno2000–Depth2	406	952	362	58	137	146

[1]Convergence criterion = 10^{-15}; SS_UPG = ssGBLUP with unknown parent groups only in **A**; SS_UPG2 = ssGBLUP with unknown parent groups for both **A** and the pedigree relationship matrix for genotyped animals.

[2]Pheno1980 = phenotypes from animals born between January 1980 and December 2018; Pheno1990 = phenotypes from January 1990 to December 2018; and Pheno2000 = phenotypes from January 2000 to December 2018. Depth3 = pedigree traced 3 generations back; Depth2 = pedigree traced 2 generations back.

and nearly unbiased when unknown parent groups correctly account for missing pedigree information for non-genotyped and genotyped animals. In this model, old generations of phenotypes and pedigree can be removed without either compromising reliability or increasing prediction biases for young selection candidates, yet reducing computing time. Additionally, this single-step GBLUP model does not require any postanalysis adjustments and can be applied to any model used in dairy genetic evaluations.

ACKNOWLEDGMENTS

This study was partially funded by Agriculture and Food Research Initiative Competitive Grant no. 2020-67015-31030 from the US Department of Agriculture's National Institute of Food and Agriculture (Washington, DC). The authors thank the Council on Dairy Cattle Breeding (Bowie, MD) for providing access to the data. The contribution of dairy producers who supplied data through their participation in the Dairy Herd Improvement program and the Dairy Records Processing Centers that edited and relayed information on to the Council of Dairy Cattle Breeding are also acknowledged. We gratefully acknowledge the editing of the manuscript by Mauricio Elzo (professor emeritus, Department of Animal Sciences, University of Florida). The authors declare they have no conflicts of interest.

REFERENCES

Aguilar, I., I. Misztal, D. L. Johnson, A. Legarra, S. Tsuruta, and T. J. Lawlor. 2010. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. J. Dairy Sci. 93:743–752. https://doi.org/10.3168/jds.2009-2730.

Bermann, M., D. Lourenco, V. Breen, R. Hawken, F. Brito Lopes, A. Legarra, and I. Misztal. 2021. Genomic evaluations with local and foreign populations. J. Anim. Sci. (accepted).

Bradford, H. L., Y. Masuda, J. B. Cole, I. Misztal, and P. M. VanRaden. 2019. Modeling pedigree accuracy and uncertain parentage in single-step genomic evaluations of simulated and US Holstein datasets. J. Dairy Sci. 102:2308–2318. https://doi.org/10.3168/jds.2018-15419.

Cesarani, A. 2021. Genomic predictions for yield traits in US Holsteins with unknown parent groups: Supplementary_Files_JDS19789. Mendeley Data V1. https://doi.org/10.17632/r3v49629jf.1.

Christensen, O. F., and M. S. Lund. 2010. Genomic prediction when some animals are not genotyped. Genet. Sel. Evol. 42:2. https://doi.org/10.1186/1297-9686-42-2.

Fragomeni, B. O., D. A. L. Lourenco, A. Legarra, P. M. VanRaden, and I. Misztal. 2019. Alternative SNP weighting for single-step genomic best linear unbiased predictor evaluation of stature in US Holsteins in the presence of selected sequence variants. J. Dairy Sci. 102:10012–10019. https://doi.org/10.3168/jds.2019-16262.

Fragomeni, B. O., D. A. L. Lourenco, S. Tsuruta, Y. Masuda, I. Aguilar, A. Legarra, T. J. Lawlor, and I. Misztal. 2015. Hot topic: Use of genomic recursions in single-step genomic best linear unbiased predictor (BLUP) with a large number of genotypes. J. Dairy Sci. 98:4090–4094. https://doi.org/10.3168/jds.2014-9125.

Gualdrón Duarte, J. L., A. S. Gori, X. Hubin, D. Lourenco, C. Charlier, I. Misztal, and T. Druet. 2020. Performances of adaptive Multi-BLUP, Bayesian regressions, and weighted-GBLUP approaches for genomic predictions in Belgian Blue beef cattle. BMC Genomics 21:545. https://doi.org/10.1186/s12864-020-06921-3.

Guarini, A. R., D. A. L. Lourenco, L. F. Brito, M. Sargolzaei, C. F. Baes, F. Miglior, S. Tsuruta, I. Misztal, and F. S. Schenkel. 2019. Use of a single-step approach for integrating foreign information into national genomic evaluation in Holstein cattle. J. Dairy Sci. 102:8175–8183. https://doi.org/10.3168/jds.2018-15819.

Howard, J. T., T. A. Rathje, C. E. Bruns, D. F. Wilson-Wells, S. D. Kachman, and M. Spangler. 2018. The impact of truncating data on the predictive ability for single-step genomic best linear unbiased prediction. J. Anim. Breed. Genet. 135:251–262. https://doi.org/10.1111/jbg.12334.

Legarra, A., I. Aguilar, and I. Misztal. 2009. A relationship matrix including full pedigree and genomic information. J. Dairy Sci. 92:4656–4663. https://doi.org/10.3168/jds.2009-2061.

Legarra, A., O. F. Christensen, Z. G. Vitezica, I. Aguilar, and I. Misztal. 2015. Ancestral relationships using metafounders: finite ancestral populations and across population relationships. Genetics 200:455–468. https://doi.org/10.1534/genetics.115.177014.

Legarra, A., and A. Reverter. 2018. Semi-parametric estimates of population accuracy and bias of predictions of breeding values and future phenotypes using the LR method. Genet. Sel. Evol. 50:53. https://doi.org/10.1186/s12711-018-0426-6.

Legarra, A., C. Robert-Granié, E. Manfredi, and J. M. Elsen. 2008. Performance of genomic selection in mice. Genetics, 180(1):611–

618. Genetics 180:611–618. https://doi.org/10.1534/genetics.108.088575.

Liu, Z., F. Reinhardt, A. Bünger, and R. Reents. 2004. Derivation and calculation of approximate reliabilities and daughter yield-deviations of a random regression test-day model for genetic evaluation of dairy cattle. J. Dairy Sci. 87:1896–1907. https://doi.org/10.3168/jds.S0022-0302(04)73348-2.

Lourenco, D. A. L., I. Misztal, S. Tsuruta, I. Aguilar, T. J. Lawlor, S. Forni, and J. I. Weller. 2014. Are evaluations on young genotyped animals benefiting from the past generations? J. Dairy Sci. 97:3930–3942. https://doi.org/10.3168/jds.2013-7769.

Lourenco, D. A. L., S. Tsuruta, B. O. Fragomeni, Y. Masuda, I. Aguilar, A. Legarra, J. K. Bertrand, T. S. Amen, L. Wang, D. W. Moser, and I. Misztal. 2015. Genetic evaluation using single-step genomic best linear unbiased predictor in American Angus. J. Anim. Sci. 93:2653–2662. https://doi.org/10.2527/jas.2014-8836.

Mäntysaari, E. A., M. Koivula, and I. Strandén. 2020. Symposium review: Single-step genomic evaluations in dairy cattle. J. Dairy Sci. 103:5314–5326. https://doi.org/10.3168/jds.2019-17754.

Masuda, Y., I. Misztal, S. Tsuruta, A. Legarra, I. Aguilar, D. A. L. Lourenco, B. O. Fragomeni, and T. J. Lawlor. 2016. Implementation of genomic recursions in single-step genomic best linear unbiased predictor for US Holsteins with a large number of genotyped animals. J. Dairy Sci. 99:1968–1974. https://doi.org/10.3168/jds.2015-10540.

Masuda, Y., P. M. VanRaden, I. Misztal, and T. J. Lawlor. 2018. Differing genetic trend estimates from traditional and genomic evaluations of genotyped animals as evidence of preselection bias in US Holsteins. J. Dairy Sci. 101:5194–5206. https://doi.org/10.3168/jds.2017-13310.

Matilainen, K., M. Koivula, I. Strandén, G. P. Aamand, and E. A. Mäntysaari. 2016. Managing genetic groups in single-step genomic evaluations applied on female fertility traits in Nordic Red Dairy cattle. Interbull Bull. 50:71–75.

Misztal, I. 2016. Is genomic selection now a mature technology? J. Anim. Breed. Genet. 133:81–82. https://doi.org/10.1111/jbg.12209.

Misztal, I., H. L. Bradford, D. A. L. Lourenco, S. Tsuruta, Y. Masuda, A. Legarra, and T. J. Lawlor. 2017. Studies on inflation of GEBV in single-step GBLUP for type. Interbull Bull. 51:38–42.

Misztal, I., A. Legarra, and I. Aguilar. 2014. Using recursion to compute the inverse of the genomic relationship matrix. J. Dairy Sci. 97:3943–3952. https://doi.org/10.3168/jds.2013-7752.

Misztal, I., S. Tsuruta, D. Lourenco, Y. Masuda, I. Aguilar, A. Legarra, and Z. Vitezica. 2018. Manual for blupf90 family of programs. Accessed on Feb. 15, 2020. http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all7.pdf.

Misztal, I., Z. G. Vitezica, A. Legarra, I. Aguilar, and A. A. Swan. 2013. Unknown-parent groups in single-step genomic evaluation. J. Anim. Breed. Genet. 130:252–258. https://doi.org/10.1111/jbg.12025.

Mrode, R. A., and G. J. T. Swanson. 2004. Calculating cow and daughter yield deviations and partitioning of genetic evaluations under a random regression model. Livest. Prod. Sci. 86:253–260. https://doi.org/10.1016/j.livprodsci.2003.09.001.

Muir, W. M. 2007. Comparison of genomic and traditional BLUP-estimated breeding value accuracy and selection response under alternative trait and genomic parameters. J. Anim. Breed. Genet. 124:342–355. https://doi.org/10.1111/j.1439-0388.2007.00700.x.

Patry, C., and V. Ducrocq. 2011. Evidence of biases in genetic evaluations due to genomic preselection in dairy cattle. J. Dairy Sci. 94:1011–1020. https://doi.org/10.3168/jds.2010-3804.

Quaas, R. L. 1988. Additive genetic model with groups and relationships. J. Dairy Sci. 71:1338–1345. https://doi.org/10.3168/jds.S0022-0302(88)79691-5.

Tsuruta, S., D. A. L. Lourenco, Y. Masuda, I. Misztal, and T. J. Lawlor. 2019. Controlling bias in genomic breeding values for young genotyped bulls. J. Dairy Sci. 102:9956–9970. https://doi.org/10.3168/jds.2019-16789.

Tsuruta, S., and I. Misztal. 2008. Computing options for genetic evaluation with a large number of genetic markers. J. Anim. Sci. 86:1514–1518. https://doi.org/10.2527/jas.2007-0324.

Tsuruta, S., I. Misztal, I. Aguilar, and T. J. Lawlor. 2011. Multiple-trait genomic evaluation of linear type traits using genomic and phenotypic data in US Holsteins. J. Dairy Sci. 94:4198–4204. https://doi.org/10.3168/jds.2011-4256.

Tsuruta, S., I. Misztal, and T. J. Lawlor. 2005. Changing definition of productive life in US Holsteins: Effect on genetic correlations. J. Dairy Sci. 88:1156–1165. https://doi.org/10.3168/jds.S0022-0302(05)72782-X.

Tsuruta, S., I. Misztal, D. A. L. Lourenco, and T. J. Lawlor. 2014. Assigning unknown parent groups to reduce bias in genomic evaluations of final score in US Holsteins. J. Dairy Sci. 97:5814–5821. https://doi.org/10.3168/jds.2013-7821.

Tsuruta, S., I. Misztal, and I. Stranden. 2001. Use of the preconditioned conjugate gradient algorithm as a generic solver for mixed-model equations in animal breeding applications. J. Anim. Sci. 79:1166–1172. https://doi.org/10.2527/2001.7951166x.

VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 91:4414–4423. https://doi.org/10.3168/jds.2007-0980.

VanRaden, P. M., M. E. Tooker, J. R. O'Connell, J. B. Cole, and D. M. Bickhart. 2017. Selecting sequence variants to improve genomic predictions for dairy cattle. Genet. Sel. Evol. 49:32. https://doi.org/10.1186/s12711-017-0307-4.

VanRaden, P. M., M. E. Tooker, J. R. Wright, C. Sun, and J. L. Hutchison. 2014. Comparison of single-trait to multi-trait national evaluations for yield, health, and fertility. J. Dairy Sci. 97:7952–7962. https://doi.org/10.3168/jds.2014-8489.

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. https://doi.org/10.3168/jds.2008-1514.

Wiggans, G. R., T. A. Cooper, P. M. VanRaden, and J. B. Cole. 2011. Adjustment of traditional cow evaluations to improve accuracy of genomic predictions. J. Dairy Sci. 94:6188–6193. https://doi.org/10.3168/jds.2011-4481.

Wiggans, G. R., and P. M. VanRaden. 1991. Method and effect of adjustment for heterogeneous variance. J. Dairy Sci. 74:4350–4357. https://doi.org/10.3168/jds.S0022-0302(91)78631-1.

Wiggans, G. R., P. M. VanRaden, and T. A. Cooper. 2012. Technical note: Adjustment of all cow evaluations for yield traits to be comparable with bull evaluations. J. Dairy Sci. 95:3444–3447. https://doi.org/10.3168/jds.2011-5000.

Xiang, T., O. F. Christensen, and A. Legarra. 2017. Technical note: Genomic evaluation for crossbred performance in a single-step approach with metafounders. J. Anim. Sci. 95:1472–1480. https://doi.org/10.2527/jas2016.1155.

ORCIDS

A. Cesarani https://orcid.org/0000-0003-4637-8669
Y. Masuda https://orcid.org/0000-0002-3428-6284
S. Tsuruta https://orcid.org/0000-0002-6897-6363
P. M. VanRaden https://orcid.org/0000-0002-9123-7278
D. Lourenco https://orcid.org/0000-0003-3140-1002
I. Misztal https://orcid.org/0000-0002-0382-1897