



Short communication: Analysis of genomic predictor population for Holstein dairy cattle in the United States—Effects of sex and age

T. A. Cooper,¹ G. R. Wiggans, and P. M. VanRaden

Animal Genomics and Improvement Laboratory, Agricultural Research Service, USDA, Beltsville, MD 20705-2350

ABSTRACT

Increased computing time for the ever-growing predictor population and linkage decay between the ancestral population and current animals have become concerns for genomic evaluation systems. The effects on reliability of US genomic evaluations from including cows and bulls in the Holstein predictor population and also from excluding older bulls from the predictor population were examined. Holstein data collected for December 2013 US genomic evaluations were used in cutoff studies to determine reliability gains, regression coefficients, and bias for 5 yield, 3 fitness, 2 fertility, and 18 conformation traits. Three predictor populations were examined based on animal sex: 30,852 cows with traditional evaluations as of August 2012, 21,883 bulls with traditional evaluations as of August 2012, and a combined group of all bulls and cows. Three subsets of the bull predictor population were examined to determine effect of age: bulls born before 1996 excluded (25% of bulls excluded), bulls born before 2001 excluded (50%), and bulls born before 2005 excluded (75%). The validation set for all predictor populations was either bulls or cows first receiving a traditional evaluation between August 2012 and December 2013. Across all traits, the addition of cows to the bull predictor population increased reliability gains by 0.4 percentage points for validation bulls and 4.4 points for validation cows. Across all traits, excluding bulls born before 1996 from the bull-only predictor population decreased gains in genomic reliability by 1.8 percentage points. For 19 of 28 traits, excluding bulls born before 2005 from the predictor population resulted in lower bias in genomic evaluations of validation bulls. Although the contribution of cows and older bulls to improved accuracy of US genomic evaluations is small, a plateau of achievable gain has not yet been reached.

Key words: single nucleotide polymorphism, genomic evaluation, dairy cattle

Short Communication

Genomic testing of cows has increased since the first unofficial USDA genomic evaluations were released in 2008 (Wiggans et al., 2011b). The number of Holstein females with genotypes included in national evaluations has increased from approximately 2,000 per month in 2010 to 12,695 per month in 2013 (Council on Dairy Cattle Breeding, 2014). In addition to increases in numbers, the female proportion of Holstein genotypes included in national evaluations has increased from 41 to 75% from 2010 to 2013.

The recent increased availability of female genotypes could provide additional genomic data for making breeding decisions. In breeds other than Holstein, which have few additional bulls with traditional evaluations that have not already been genotyped, the primary opportunity for new data to improve genomic predictions is from females. In a simulation, Thomasen et al. (2014) showed that a small population can benefit from the addition of cow information. However, many countries with genomic evaluations do not include cows in their predictor populations (Reinhardt et al., 2009; Schenkel et al., 2009; Spelman et al., 2010) because of concern about bias in cow evaluations and complexity involved in using them.

Including a cow's own yield records in genomic predictions leads to an overestimated genomic evaluation (Dassonneville et al., 2012). To incorporate traditional evaluations of cows into the US genomic evaluation system and increase overall gains in reliability, 2 adjustments were made to traditional yield data. The first adjustment reduced the mean and variance of traditional evaluations for genotyped cows to be similar to those for bulls (Wiggans et al., 2011a). To restore comparability of traditional evaluations for genotyped and nongenotyped cows, a second adjustment was applied to all cows by reducing the range of the deregressed Mendelian sampling within birth-year group (Wiggans et al., 2012). Beginning with the December 2014 evaluations, the heritability was lowered, which eliminated the need for this second adjustment.

The number of bulls and cows with traditional information in the predictor population is also growing.

Received September 24, 2014.

Accepted December 18, 2014.

¹Corresponding author: Tabatha.Cooper@ars.usda.gov

As of December 2013, the majority of Holstein bulls in the current US predictor population was born between 2001 and 2010, but 9,540 bulls that were born before 2001 contributed to the US genomic evaluation system. Increased computing time for the ever-growing predictor population and linkage decay between the ancestral population and current animals have become concerns. A simulation study by Lourenco et al. (2014) suggests that truncating old data does not reduce accuracy of genomic prediction.

Objective of Study

The objective of this study was to determine the effects on reliability of US genomic PTA from including cows and bulls in the Holstein predictor population and also from excluding older bulls from the predictor population.

Holstein data collected for December 2013 US genomic evaluations were provided by the Council on Dairy Cattle Breeding (Reynoldsburg, OH) and used in a cut-off study (VanRaden et al., 2009) to determine reliability gains due to the inclusion or exclusion of genomic information from the predictor population to calculate SNP effects based on animal sex and age. Only animals with both a genotype and a traditional evaluation were included. Three predictor populations were examined based on animal sex: 30,852 cows with traditional evaluations as of August 2012, 21,883 bulls with traditional evaluations as of August 2012, and a combined group of all bulls and cows. Three subsets of the bull predictor population were examined to determine effect of age: bulls born before 1996 excluded (17,047 remaining), bulls born before 2001 excluded (11,507 remaining), and bulls born before 2005 excluded (6,623 remaining). The number of bulls excluded from each subset was equivalent to removal of 25, 50, and 75%, respectively, of the total bull predictor population. The validation set for all predictor populations was either bulls or cows that received their first traditional evaluation between August 2012 and December 2013. Validation bulls were required to have ≥ 10 daughters, and validation cows were required to have been genotyped before 2 yr of age to avoid selection bias that may have occurred if the cow was selected to be genotyped because of her own lactation performance. The number of animals in each validation set varied by trait and ranged from 337 to 1,486 for bulls and from 358 to 26,559 for cows, reflecting the varying amount of information available on each animal. For example, productive life records may take many years to accumulate, whereas production trait information is available much sooner.

Reliability gains for the validation set were calculated as the reliability of genomic PTA, which included

SNP and polygenic effects estimated from August 2012 predictor populations (genotyped animals with traditional evaluations) and August 2012 traditional parent averages, minus parent-average reliability. Additional statistics included coefficients for regression of December 2013 daughter deviations on August 2012 genomic PTA and bias calculated as December 2013 daughter deviation minus August 2012 genomic PTA and are presented in Supplemental Tables S1 to S9 (<http://dx.doi.org/10.3168/jds.2014-8894>).

Evaluations for 28 traits were examined: yield traits (milk, fat, and protein yields, and fat and protein percentages); functional traits (productive life, SCS, daughter pregnancy); fertility traits (heifer conception and cow conception rates); and conformation traits [final score, stature, strength, dairy form, body depth, rump angle and width, rear legs (side and rear views), feet and legs, foot angle, fore udder attachment, rear udder height, udder depth and cleft, front and rear teat placements, and teat length]. Net merit was not examined because the reported value is computed as an index of the genomic PTA for individual traits and not as a single trait. The number of years between the predictor population and validation set was decreased, 2.5 yr compared with the usual 4 yr, to allow for the inclusion of cow information. Overall gains in reliability, regression coefficients, and bias can be skewed due to the shorter interval; however, comparisons between predictor population groups are valid.

Gains in Reliability from Genomics

Mean reliability gains over parent average for validation cows across all traits (Table 1) were 20.0 percentage points for the cow predictor population, 21.0 percentage points for the bull predictor population, and 25.4 percentage points for the combined predictor population. Mean reliability gains over parent average for validation bulls across all traits were 19.3 percentage points for the cow predictor population, 32.6 percentage points for the bull predictor population, and 33.0 percentage points for the combined predictor population. Across all traits, the inclusion of cow information increased genomic reliability by 0.4 percentage points for the validation of bulls and 4.4 percentage points for the validation of cows. The small gain in genomic reliability by using the combined predictor population is consistent with the accuracy results of Calus et al. (2013).

Mean reliability gains over parent average for validation bulls using subsets of the bull-only predictor population based on bull age (Table 2) were 30.8, 28.1, and 20.2 percentage points when bulls born before 1996, 2001, and 2005, respectively, were excluded; mean gain

Table 1. Reliabilities for August 2012 traditional parent averages of validation bulls and cows and gains in reliability for genomic PTA compared with parent average by trait and predictor population¹

Trait group	Parent average reliability (%)		Reliability gain (percentage points) ²					
	Validation cows	Validation bulls	Validation cows			Validation bulls		
			Predictor cows	Predictor bulls	Predictor cows and bulls	Predictor cows	Predictor bulls	Predictor cows and bulls
Yield ³	26.1	40.5	34.4	41.4	42.6	31.3	40.5	40.9
Functional ⁴	22.8	35.9	2.7	10.9	13.3	11.1	36.2	36.3
Fertility ⁵	24.5	28.4	13.5	9.0	14.2	8.2	42.6	41.3
Conformation ⁶	23.4	37.1	19.6	18.3	23.9	18.5	28.7	29.3
Overall	23.9	37.0	20.0	21.0	25.4	19.3	32.6	33.0

¹SNP and polygenic effects estimated from cows only, bulls only, or bulls and cows with traditional PTA by August 2012.

²Genomic reliability minus parent-average reliability.

³Yield = milk, fat, and protein yields, and fat and protein percentages.

⁴Functional = productive life, SCS, and daughter pregnancy rate.

⁵Fertility = heifer and cow conception rates.

⁶Conformation = final score, stature, strength, dairy form, body depth, rump angle and width, rear legs (side and rear views), feet and legs, foot angle, fore udder attachment, rear udder height, udder depth and cleft, front and rear teat placements, and teat length.

for all bulls was 32.6 points. Across traits, excluding bulls born before 1996 decreased gains in genomic reliability by 1.8 percentage points compared with gains for all bulls. For yield traits, excluding bulls born before 1996 increased reliability gains for fat and protein yields and protein percentage. Compared with Lourenco et al. (2014), this study shows an overall decrease in reliability gains when old bulls are removed from the predictor set in contrast to having no effect. However, when individual trait groups are examined, an increase was found for some.

To compare the effect of excluding the same numbers of bulls randomly rather than by bull age, a separate cutoff study was performed for milk yield only. Gains in

genomic reliability for PTA milk were 34.9 percentage points for all bulls and 33.5, 28.5, and 19.9 percentages points with random exclusion of 25, 50, and 75% of bulls, respectively. The decreases were much greater than for the comparable reliability gains of 34.7, 33.5, and 30.2 percentage points when bulls were excluded from the predictor population based on birth year.

For Holsteins, which have a large number of bulls with high-reliability PTA, including cows in addition to bulls in the predictor population contributed only slightly to increasing the reliability of genomic PTA. However, a cow-only predictor population still provided an increase in genomic reliability compared with parent-average reliability. If the number of bull genotypes

Table 2. Reliability gains for genomic PTA compared with parent average from August 2012 traditional parent averages for validation bulls¹ by trait and age of bulls in bull predictor population²

Trait group	Reliability gain (percentage points) ³			
	All bulls	Birth year ≥ 1996	Birth year ≥ 2001	Birth year ≥ 2005
Yield ⁴	40.5	40.5	38.8	35.2
Functional ⁵	36.2	34.7	31.7	21.2
Fertility ⁶	42.6	29.6	26.8	8.2
Conformation ⁷	28.7	27.6	24.7	17.2
Overall	32.6	30.8	28.1	20.2

¹Received traditional PTA between August 2012 and December 2013 and had ≥10 daughters.

²SNP and polygenic effects estimated from bulls with traditional PTA by August 2012.

³Genomic reliability minus parent-average reliability.

⁴Yield = milk, fat, and protein yields, and fat and protein percentages.

⁵Functional = productive life, SCS, and daughter pregnancy rate.

⁶Fertility = heifer and cow conception rates.

⁷Conformation = final score, stature, strength, dairy form, body depth, rump angle and width, rear legs (side and rear views), feet and legs, foot angle, fore udder attachment, rear udder height, udder depth and cleft, front and rear teat placements, and teat length.

is limited, cow information becomes more valuable. Older bulls also contributed only slightly to increasing genomic reliability because of linkage decay between the validation and ancestral populations. For consistency across trait and breed, evaluations will continue to include both cow and bull information. Although the contribution of cows and older bulls to improved accuracy of US genomic evaluations is small, a plateau of achievable gain has not yet been reached.

ACKNOWLEDGMENTS

The cooperation of the Council on Dairy Cattle Breeding (Reynoldsburg, OH) in supplying pedigree, performance, and genotypic data is acknowledged.

REFERENCES

- Calus, M. P. L., Y. de Haas, and R. F. Veerkamp. 2013. Combining cow and bull reference populations to increase accuracy of genomic prediction and genome-wide association studies. *J. Dairy Sci.* 96:6703–6715.
- Council on Dairy Cattle Breeding. 2014. Genotypes included in evaluations by breed, chip density, presence of phenotypes (old vs. young), and evaluation year-month (cumulative). Accessed Sep. 22, 2014. https://www.cdcb.us/Genotype/cur_density.html.
- Dassonneville, R., A. Baur, S. Fritz, D. Boichard, and V. Ducrocq. 2012. Inclusion of cow records in genomic evaluations and impact on bias due to preferential treatment. *Genet. Sel. Evol.* 44:40.
- 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.
- Reinhardt, F., Z. Liu, F. Seefried, and G. Thaller. 2009. Implementation of genomic evaluation in German Holsteins. *Interbull Bull.* 40:219–226.
- Schenkel, F. S., M. Sargolzaei, G. Kistemaker, G. B. Jansen, P. Sullivan, B. J. Van Doormaal, P. M. VanRaden, and G. R. Wiggans. 2009. Reliability of genomic evaluation of Holstein cattle in Canada. *Interbull Bull.* 39:51–58.
- Spelman, R. J., J. Arias, M. D. Keehan, V. Obolonkin, A. M. Winkelman, D. L. Johnson, and B. L. Harris. 2010. Application of genomic selection in the New Zealand dairy cattle industry. Proc. 9th World Congr. Genet. Appl. Livest. Prod., Leipzig, Germany, Commun. 0311. Gesellschaft für Tierzuchtwissenschaften e.V., Gießen, Germany. Accessed Sep. 17, 2014. <http://www.kongressband.de/wcgalp2010/assets/pdf/0311.pdf>.
- Thomasen, J. R., A. C. Sørensen, M. S. Lund, and B. Guldbandsen. 2014. Adding cows to the reference population makes a small dairy population competitive. *J. Dairy Sci.* 97:5822–5832.
- 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.
- Wiggans, G. R., T. A. Cooper, P. M. VanRaden, and J. B. Cole. 2011a. Technical note: Adjustment of traditional cow evaluations to improve accuracy of genomic predictions. *J. Dairy Sci.* 94:6188–6193.
- Wiggans, G. R., P. M. VanRaden, and T. A. Cooper. 2011b. The genomic evaluation system in the United States: Past, present, future. *J. Dairy Sci.* 94:3202–3211.
- 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.