Efficiency of including cows in genomic prediction versus multivariate models for fertility traits in dairy cattle in the United Kingdom

R. Mrode and M. Coffey
Scotland Rural College, Roslin Institute Building, Easter Bush, EH25 9RG, Midlothian, UK

ABSTRACT: The efficiency of genomic evaluations with including cows in the reference population was compared to multivariate genomic evaluations for calving interval (CI) and non-return rate at 56 days (NR56). Cow data included constituted 10 and 15% of the reference set for CI and NR56 respectively. Analyses were based on SNP BLUP, BayesA and BayesC models. Inclusion of cow data had very little impact on accuracies and errors of predictions of genomic evaluations for CI and NR56. The bivariate genomic models resulted in an increase of 15% in accuracy for NR56. Similar increases in accuracies of 10 to 15% were observed for NR56 from the Bayesian models. All of the analyses had very little effect on accuracies for CI. The study concluded that multivariate SNP BLUP models resulted in more efficient genomic evaluations rather than including cows.
Keywords: Multivariate genomic models; Bayesian methods; Genomic evaluations; Fertility traits

Introduction

Genomic prediction and selection on the basis of single nucleotide polymorphisms (SNP) from dense marker maps in practical animal breeding has not only grown in popularity in recent years but is becoming the method of choice. However, most practical applications in the dairy sector are still based on univariate models. This is partly to avoid huge computational requirements due to the large number of bulls in the reference populations. However, some countries have considered including females in their reference population in an attempt to further boost the accuracy of genomic prediction (Wiggans et al. 2012 ; Cooper et al 2014). There is however the issue of bias in cow evaluations although this may not be of much concern in non-production traits. However, input variables for cows into genomic predictions in the form of de-regressed proofs or yield deviations are usually of lower accuracy and will tend to limit their contribution to the accuracy of genomic prediction especially for traits of low heritability. Given that the use of SNP BLUP models with residual updating (Legarra and Misztal, 2008) is less demanding in computational cost, the question is whether a multi-trait application for traits of low heritabilities which are conventionally analysed with a multi-trait model might be more beneficial than including females in reference populations per se? Alternatively, multi–trait models with females also included in the reference population could result in further increases in accuracy of genomic prediction if it is computationally

feasible. Currently, single trait SNP BLUP is implemented for calving interval (CI) and non-return rate at 56 days (NR56) although conventional evaluations are from a multi-trait model in the United Kingdom (UK). This paper examines the accuracy of genomic predictions for CI and NR56 from a single trait SNP BLUP, BayesA and BayesC models with and without females included in the reference population. Secondly, accuracies were evaluated from multivariate SNP BLUP models for both traits with and without females in the reference population.

Materials and Methods

Data. Data from 13893 bulls and 1514 cows with 50k genotypes from Illumina 50k chip and deregressed proofs for CI and NR56 were available for the analysis. The deregressed proofs used were from the UK official December 2013 run and MACE proofs. The software MiX99 (Lidauer, et al, 2011) was used for the deregression using a full animal pedigree with effective daughter contributions (EDCs) used as weights. The evaluations for cows were scaled to have the same mean and variance as bulls prior to the deregression (Wiggans et al , 2011) and EDCs for cows were computed from cow reliabilities after removing parental contributions. The details of the various edits carried on the SNP data were as described by Mrode et al, (2011). The total number of SNPs selected for genomic evaluations after the various edits was 43121.

Statistical analyses. The statistical model used for the estimation of SNP effects is:

$$y_i = \mu + v_i + \sum_{j=1}^{m} Z_{ij} u_i + e_{ij};$$

where y_i is the deregressed proof of a bull, μ is the overall mean, v_i is the residual polygenic effect (20% of additive genetic variance) of ith bull, $\mathbf{Z}_{ij}$ is the genotype value coded as 0, and 2 for the homozygotes and 1 for the heterozygote, u_i is the random regression coefficient for j[th] SNP and e_i is the residual effect. Bulls born before 2006 and all cows were used as the reference population and were used to estimate the SNP effects for fertility traits. Bulls born in 2006 and after were used for the purposes of validation. Any bull in the validation set who was a sire of cows in the reference population was excluded. The bulls in both data sets were required to have at least 8 EDCs and a reliability of at least 55% for CI and 40% for NR56. There were 9676 and 10257 bulls in the reference sets for CI and NR56 respectively. Corresponding values with cows in-

cluded were 10773 and 11771 respectively. The bulls in the validation sets were 2688 for CI and 3636 for NR56 in all analyses. The bivariate analyses allowed for missing values, thus there were 11522 bulls in the reference population and 4042 bulls in the validation set. The inclusion of cows in the bivariate analysis resulted in a reference population of 13194 animals. The software MiX99 was used for all SNP BLUP analysis while modified Club ware software (Mrode et al, 2010) was used for all Bayesian analyses. The Bayesian models did not include any polygenic effects in the model. The Gibbs sampling chain was run for 40000 times for the Bayesian methods with the 12000 discarded as burn-in period. The proportion of SNPs with zero effects assumed for BayesC was 30%. The genetic parameters used for SNP BLUP analyses were derived from conventional estimates with a heritability of 0.033 for CI and 0.019 for NR56, genetic and residual correlations of -0.449 and -0.341 respectively between both traits. The accuracy of evaluations were computed as correlation of direct genomic values with de-regressed proofs in the validation set while differences between both variables were used to estimate errors of prediction.

Results and Discussion

A summary of the univariate animals with or without cows in the reference population is given in Table 1. In this study the percentage of cows included in the reference population was about 10% for CI and 13% for NR56 and this seems to have no effect on the accuracy of evaluations or error of prediction for both traits. Indeed the number of cows included in the reference population in this study is rather limited due to the limited genotyping of cows at this time in the UK. However, the results from Cooper et al, (2014) where the proportion of cows was 59% in the reference population (reference population of 52735 animals), the inclusion of cows only resulted in an increase of 0.5% for production traits and resulted in reduced accuracy for fitness traits such as heifer conception and daughter pregnancy rate. It seems that the usual low reliability of cow evaluations implies they contribute little or no information when a reasonable number of bulls is already in the reference population.

Table 1. Summary of accuracies of evaluations from univariate genomic evaluations with or without cows in the reference population

| | SNP BLUP | | | |
| | Bulls only | | Bulls and Cows | |
	CI	NR56	CI	NR56
Reg.	1.011	1.138	1.004	1.134
Corr.	0.702	0.332	0.702	0.333
MD.	0.225	-0.675	0.212	-0.729
MS.	2.927	4.745	2.926	4.754
	BayesA			
Reg.	0.878	1.023	0.844	1.106
Corr.	0.690	0.358	0.606	0.344
MD	0.162	-0.701	-0.405	-0.316
MS	3.060	4.703	3.385	4.732
	BayesC			
Reg.	0.848	0.988	0.848	1.010
Corr.	0.712	0.384	0.712	0.383
MD	0.252	-0.672	0.259	-0.600
MS	3.009	4.650	3.010	4.653

Reg. = regression coefficients; Corr. = correlations, MD and MS are means and standard deviations of differences between de-regressed proofs and direct genomic values

Table 2. Summary of accuracies of evaluations from bivariate genomic evaluations with or without cows in the reference population

| | SNP BLUP | | | |
| | Bulls only | | Bulls and Cows | |
	CI	NR56	CI	NR56
Reg.	1.009	1.077	1.003	1.071
Corr.	0.718	0.382	0.718	0.382
MD.	-1.376	-0.0758	-1.373	0.360
MS.	2.947	3.880	2.946	3.894

Reg. = regression coefficients; Corr. = correlations, MD and MS are means and standard deviations of differences between de-regressed proofs and direct genomic values

The results indicated that NR56 with a very low heritability of about 0.019 benefitted substantially in accuracy from the application of Bayesian methods, with a gain of 8% from BayesA and 15% from BayesC.

The results of the bivariate analyses are summarized in Table 2. Similar to the univariate analyses, the inclusion of cows in the reference population had no effect on estimates of accuracies. The accuracies of CI from the bivariate models with or without females in the reference population are not different from corresponding estimates from the univariate models. However, the accuracy of NR56 increased by about 15% from the bivariate SNP BLUP model without females in the reference population relative to results from the univariate models. This is consistent with the observation that traits with lower heritabilities benefit more when analyzed with traits with higher heritabilities (Thompson and Meyer, 1986). This increase in accuracy for NR56 is similar to that observed from a univariate BayesC model. The mean differences indicated that multivariate analyses resulted in a slightly over prediction for CI.

Conclusion

The inclusion of cows in the reference population did not have any impact on accuracies of genomic evaluations for CI and NR56. There was a gain in accuracy of about 15% in NR56 from the multivariate model even without females in the reference population relative to the univariate models. A similar gain in accuracy for NR56 was also observed from the univariate BayesC model. However computing time for the latter is longer; therefore the multi-

trait model seems a better option to increase the accuracy of low heritability traits rather than inclusion of cow genotypes.

Literature Cited

Cooper, T. A., Wiggans, G. R. and VanRaden, P. M. (2014) Proc. 10th World Congr. Genet. Appl. Livest. Prod. , Vanvouver, Canada (Submitted).

Legara, A. and Misztal, I. (2008). J. Dairy Sci., 91:360-366

Lidauer, M., Matilainen, K, Mäntysaari, E. et al. (2011). MiX99 Manual. MTT, Jokioinen, Finland

Mrode, R., Coffey, M., Strandén, I. et al. (2010). Proc 9th World Congr. Genet. Appl. Livest. Prod. 0518

Mrode, R., Krzyzelewski, T., Moore, K. et al. (2011) Interbull Bulletin.44:173-174

Thompson, R. and Meyer, K. (1986). Livest. Prod. Sci. 15:299-313

Wiggans, G.R., Cooper T.A., VanRaden P.M. et al. (2011) J. Dairy Sci. 94:6188-6193.

Wiggans, G.R., VanRaden, P.M., and Cooper T.A. (2012) J. Dairy Sci. 95:3444-3447.