The Impact of Residual Polygenic Effect on Genomic Evaluation

Z. Liu, F. Seefried, F. Reinhardt and R. Reents
vit *w.V., Heideweg 1, 27283 Verden, Germany*

Abstract

Genomic evaluation has being developed or already implemented for dairy cattle breeding in an increasing number of countries. Genotyped young calves without own progeny have been compared to usually older bulls tested via conventional progeny testing programme. Concerns have been raised in several countries that genomic pedigree index (GPI) of young candidates may have too high variance, which may lead to a problem in correctly ranking animals and in stability of GPI over time. One solution to this problem may be to include a residual polygenic effect (RPG) or to increase the variance of RPG in the estimation of SNP effects or direct genomic values (DGV). Three levels of RPG variance, 5%, 10% and 20%, of total genetic variance were compared to the current SNP model for German Holsteins with a very low heritability of 0.0001 for RPG. MACE or national proofs of 44 traits and EuroGenomics genotypes from April 2010 genomic evaluation were used and a validation study was conducted for each of the three scenarios following Interbull GEBV validation test procedure. Though variance of estimated SNP effects dropped significantly with an increasing RPG variance, correlation of SNP effect estimates were very high between any pair of the scenarios. For training bulls, correlation of DGV with conventional EBV or DGV variance decreased with increasing RPG variance, however, correlation of GEBV with conventional EBV or variance of GEBV changed only marginally. It appeared that the loss of DGV in correlation with EBV and variance was compensated by the RPG effect. For validation bulls, correlation of DGV or GPI with deregressed proof (DPRF) dropped slightly with increasing RPG variance for some traits, but remained at a similar level for the others. Regression slop estimates of DPRF on validation bulls' GPI exceeded 1 with a higher RPG variance for production or somatic cell scores traits, indicating the assumed RPG variance may be too high for those traits. However, the regression slope estimates for some conformation traits indicated 5% or 10% RPG being optimal. Optimal variance of RPG appeared to vary across the analysed traits. For those traits with high heritability or reliability, lower percentage of RPG variance seemed to be appropriate.

1. Introduction

Genomic evaluation system has being developed or already implemented for routine genomic selection of dairy cattle in many countries (Liu *et al.*, 2010; VanRaden, 2008). The genomic evaluation system is usually validated using a data truncation technique like Interbull's GEBV test (Mäntysaari *et al.*, 2010), in which SNP effects or DGV are estimated based on an earlier data set, simulating a genomic evaluation four years ago, and then applied to a group of validation bulls for comparing their GPI to DPRF from a current conventional evaluation. Regression of DPRF on GPI must not deviate significantly from its expectation, if the genomic evaluation system is unbiased. However, several countries have seen over scaled GPI for young calves, i.e. GPI having too high variance, through the validation test or examining the proportion of genotyped calves in top ranking lists, among other statistical procedures. As genomic reference population becomes larger and more complete, i.e. more candidates having sires genotyped, a likely double counting of pedigree index may happen when candidates' DGV are combined later with conventional PI. A solution to the over scaled GPI or double counting of PI may be to fit a residual polygenic effect or to increase the variance of RPG in genomic evaluation model. The objective of this study was to investigate the impact of RPG on genomic evaluation via a genomic validation.

2. Materials and Methods

2.1. A genomic evaluation model materials

A statistical model was applied to genotypic and phenotypic data of reference bulls (Reinhardt *et al.,* 2009):

$$q_i = \mu + v_i + \sum_{j=1}^{p} z_{ij} u_j + e_i$$

where q_i is DPRF of bull *i*, μ is a general mean, v_i is RPG of bull *i*, *p* is the number of fitted SNP markers ($j = 1, \cdots, 45181$), z_{ij} is genotype value of marker *j* of bull *i*, u_j is random regression coefficient for marker *j*, and e_i is residual effect for the record of bull *i*. Until now, a very low heritability is set to 0.0001 for the RPG in German Holstein genomic evaluation. Fitting the RPG in the SNP model can account for the fact that the markers may not explain all genetic variance and it may also avoid the problem that the markers capture the relationship among animals if the genomic model does not include the polygenic effect. The fitted RPG of the genomic models was analysed in the same way as in conventional genetic evaluation, i.e. using full pedigree and identical grouping of phantom parent groups. As a direct estimation of the RPG variance under the SNP model with REML is not feasible yet for large reference population at present time, three levels of RPG variance were investigated (Table 1).

Table 1. Scenarios for variance of RPG.

	Current SNP model	Proportion of residual polygenic variance		
		A	B	C
Scenario	$h^2_{RPG} = .0001$	5%	10%	20%

2.2. Data materials

MACE or national bull proofs from April 2010 conventional evaluations were used as phenotypic data. Genotype data from the April 2010 German Holstein genomic evaluation, including EuroGenomics bulls, were considered for this study. The number of genotyped animals amounted to 26,191, and the number of genotyped Holstein reference bulls with daughters in milk was 17,429. For conducting a validation study, three groups of traits were distinguished: early measured traits (non-return rate and interval first to successful insemination heifer, stillbirth and calving ease direct), late measured traits (longevity, interval first to successful insemination cow and days open), and all remaining 37 traits (Liu *et al.,* 2010). Table 2 shows genomic reference populations for both full data set and reduced subset for validation.

Table 2. Genomic reference populations for full data set and reduced subset for validation.

Birth year	Milk yield		Longevity		NR heifer	
	Full	Sub-set	Full	Sub-set	Full	Sub-set
≤ 1995	4552	4552	3790	3784	2893	2892
1996	774	774	656	656	525	525
1997	738	738	687	687	486	486
1998	1195	1195	1055	1055	919	919
1999	1419	1419	1077	1077	1228	1228
2000	1519	1519	1045	1045	1354	1354
2001	1540	1540	1057	1057	1262	1262
2002	1583	1583	963		1298	1298
2003	1972	1174	1109		1744	1744
2004	1721		193		1601	
2005	416				752	
2006					220	
Sum	17429	14494	11632	9361	14282	11708

3. Results

3.1. SNP effect estimates

For a total number of 44 traits, SNP and RPG effects were estimated using a reduced genomic reference population for each of the three scenarios. Since the RPG effect was analysed in the same way as in conventional evaluation, many more animals were considered for this effect than DGV. For example, 28318 ancestors were identified via sire-dam relationship for 14494 genotyped bulls in reference population, together with 98 phantom parent groups. In contrast, DGV effect was fitted only for the genotyped bulls. Estimated SNP effects from all the scenarios were compared (Table 3) for milk yield. With an increasing difference in RPG variance between two runs, correlation of SNP effect

estimates decreased slightly. But the level of the correlations remained very high. As the RPG variance dropped, variance of SNP effect estimates, relative to the current model, decreased and so did the estimate of the largest SNP effect. Similar findings were obtained for all the other traits as well.

Table 3. Variance and correlation of SNP effect estimates between the scenarios (milk yield).

	SNP effect variance	Largest SNP effect	Correlations of SNP effects of two scenarios		
			A	B	C
$h^2_{RPG} = .0001$	1	1	.942	.910	.860
A (5%)	.65	.84		.993	.964
B (10%)	.50	.75			.987
C (20%)	.34	.62			

3.2. Estimated DGV and GEBV

For the training bulls, sum of DGV and RPG was defined as GEBV. Table 4 shows variance of estimated DGV or GEBV and correlations with conventional EBV. It can be seen that the DGV correlation with EBV decreased with increasing RPG and GEBV correlation increased slightly. Relative to the variance of EBV, DGV variance dropped significantly as RPG variance increased. However, the variance of GEBV remained the same, indicating that the loss of DGV was compensated by RPG for the training bulls. For all the scenarios, regressions of conventional EBV or DPRF on DGV and RPG estimates were 1 for the reference bulls, and the regression intercepts were almost zero. RPG and DGV estimates were found to be positively correlated for milk yield, the higher the percent of RPG variance, the higher the correlation value.

It can be seen in Table 5 that GEBV estimates of milk yield were essentially equal for the training bulls between any pair of the scenarios. The DGV correlations between the scenarios were close to unity, except those of the current model. Despite the nearly unity correlation of DGV estimates between the scenarios, variance of DGV estimates differed significantly.

Table 4. Variance and correlation of DGV or GEBV with conventional EBV for bulls in reference population (milk yield).

DGV (first) GEBV (second row)	Correlation with EBV	Variance divided by EBV variance
$h^2_{RPG} = .0001$	.945 .945	.95 .96
A (5%)	.895 .964	.57 .94
B (10%)	.873 .973	.47 .95
C (20%)	.842 .984	.36 .96

Table 5. Variance and correlation of DGV or GEBV between scenarios for the training bulls (milk yield).

DGV (first) GEBV (second row)	Relative overall variance	Correlation between two scenarios		
		A	B	C
$h^2_{RPG} = .0001$	1 1	.954 .994	.935 .990	.909 .981
A (5%)	.60 .99		.997 .999	.984 .994
B (10%)	.50 1			.995 .998
C (20%)	.38 1.01			

3.3. Correlation of DGV or GPI with DPRF for validation bulls

Following Interbull GEBV test procedure, conventional DPRF of the validation bulls were compared to their DGV or GPI estimates calculated based on the reduced subset of the reference population. Table 6 shows Pearson correlation of DPRF, without adjusting for reliability of DPRF, with the bulls' DGV estimates in first row and GEBV in second row for a selected group of traits. These correlations remained at a high level, indicating a high reliability of the genomic evaluation system. The DGV correlations with DPRF decreased, especially for milk yield, as the variance of RPG increased. In comparison, the GEBV correlations reduced less or remained constant, e.g. for SCS.

Table 6. Pearson correlations between DGV or GPI and DPRF for validation bulls.

DGV (first) GPI (second row)	$h^2_{RPG} = .0001$ ‡	Percent of residual polygenic variance		
		5%	10%	20%
Milk yield	.76	.73	.71	.70
	.76	.75	.74	.74
SCS	.72	.71	.70	.68
	.72	.73	.72	.72
Stature	.73	.73	.72	.70
	.72	.71	.71	.71
Udder depth	.72	.71	.70	.68
	.70	.70	.69	.68
BCS	.62	.62	.62	.61
	.61	.58	.58	.58

‡ based on January 2010 data.

3.4. Regression of deregressed proofs on GPI for validation bulls

Regression of conventional DPRF of the validation bulls on their GPI based on earlier phenotypic information can identify possible bias of a genomic evaluation model. Considering standard error of its estimate, regression intercept was not significantly different from 0 for all the traits. Regression slop estimate lower or higher than its expected value, nearly unity for the validation population, indicates that variance of GPI being too high or too low, respectively. For traits with high heritability or reliability, e.g. production traits, SCS, stature and rump angle, the optimal RPG variance appeared to be 5% or less (Table 7). For the conformation traits rump width and BCS, 10% or higher RPG variance gave less biased GPI estimates.

Table 7. Regression slop estimates of DPRF on GPI for validation bulls.

Trait	$h^2_{RPG} = .0001$ ‡	Percent of residual polygenic variance		
		5%	10%	20%
Milk, kg	.93	1.17	1.26	1.40
Fat, kg	.96	1.15	1.24	1.38
Protein,kg	.89	1.13	1.23	1.37
SCS	.97	1.13	1.21	1.34
Longevity	.97	.83	.90	1.00
Stature	.91	1.00	1.09	1.21
Rump an.	.96	1.05	1.12	1.22
Rump w.	.83	.84	.89	.97
Udder dp.	1.01	1.19	1.26	1.36
BCS	.95	.94	1.00	1.09
M. speed	1.01	1.06	1.11	1.19

‡ based on January 2010 data.

4. Discussion

In order to solve the problem of over scaled GPI for genotyped calves in genomic evaluation, three levels of RPG variance were studied and compared to the current SNP model having nearly zero RPG variance. The three models with varying RPG variance were applied to the genotypes and conventional proofs obtained from the April 2010 German Holstein genomic evaluation. A validation study was conducted following Interbull GEBV test procedure. SNP effect estimates were very highly correlated between the scenarios, however, variance of SNP effect estimates reduced significantly with increasing

RPG variance. Regardless of the RPG variance, GEBV of the training bulls were essentially equal between any two scenarios. In contrast to GEBV, DGV variance dropped more than DGV correlation with EBV. The loss in DGV was compensated by the RPG effect for the training bulls. For the validation bulls correlation of GPI with conventional DPRF reduced slightly with increasing RPG variance, though the reduction was less for GPI than DGV. Optimal regression slope of DPRF on GPI seemed to vary across the traits. RPG variance of 5% or less appeared to be appropriate for traits with high heritability or reliability. However, for some conformation traits, the optimal value for RPG variance should be 10% or higher. Fitting a RPG effect in SNP effect estimation affected the correlation of GPI with DPRF less than variance of GPI. The impact of the fitted RPG was higher for estimated DGV than GPI or GEBV.

5. References

Ducrocq, V. & Liu, Z. 2009. Combining genomic and classical information in national BLUP evaluations. *Interbull Bulletin 40,* 172-177.

Liu, Z., Seefried, F., Reinhardt, F. & Reents, R. 2010. Approximating reliabilities of estimated direct genomic values. *Interbull Bulletin 41,* 29-32.

Mäntysaari, E., Liu, Z. & VanRaden, P. 2010. Interbull validation test for genomic evaluations. *Interbull Bulletin 41,* 17-21.

Reinhardt, F., Liu, Z., Seefried, F. & Thaller, G. 2009. Implementation of genomic evaluation in German Holsteins. *Interbull Bulletin 40,* 219-226.

VanRaden, P.M. 2008. Efficient methods to compute genomic predictions. *J. Dairy Sci. 91,* 4414-4423.