Data Selection and Pilot Run on
Simplified Genomic MACE (S-GMACE)

B. Zumbach, J. Jakobsen, F. Forabosco, H. Jorjani and J. Dürr

Interbull Centre, Box 7023, SE-750 07 Uppsala, Sweden

Abstract

As a first step to international evaluation of genomically enhanced breeding values (GEBV) a pilot run on Simplified Genomic Multiple Across Country Evaluation (S-GMACE) was performed. This implies that only one GEBV per bull is considered in the analysis. GEBV data for Holstein bulls corresponding to the December 2010 routine MACE run from 3 to 7 countries for the traits protein yield, somatic cell score, stature, direct longevity and cow conception 1 (measures as conception rate) were available. For bulls with several GEBVs only one was chosen based on the following hierarchical selection criteria: largest number of daughters, country of first registration, highest correlation to country of first registration, and highest genomic reliability. The GEBV data were merged with the corresponding pre-edited conventional EBVs. In cases where both a GEBV and an EBV for the same bull existed within a country the EBV was discarded. The final dataset ranged from 68,481 (cow conception 1) to 151,904 records (protein yield) with a GEBV proportion from 40% (protein yield) to 52% (cow conception 1). The data were analyzed with MACE methodology using sire-dam relationships. Genetic trends of conventional and genomic runs coincided. Young bulls were among the top 100. In the case of protein yield they ranged from 26 to 64 according to the country scale. Due to the inclusion of genomic information there was an overall increase in international reliability on all country scales. Harmonization of input data is important in order to fairly select among GEBVs from different sources.

Introduction

Genomic evaluations in dairy cattle have been introduced in a number of countries. An international evaluation taking into account genomically enhanced breeding values (**GEBV**) is not yet in place. While residuals are assumed to be uncorrelated in traditional MACE (Multiple Across Country Evaluation, see Schaeffer, 1994), residual correlations in genomic MACE (**GMACE**) are nonzero. Methodology and software to handle this problem have been developed (Sullivan and VanRaden, 2010; VanRaden and Sullivan, 2010). However, it requires the information on the proportion of shared genotypes among countries. As this information is currently not available a simplified genomic MACE (**S-GMACE**) considering only one GEBV per bull has been suggested. In the presence of multiple GEBVs for one bull, only one GEBV will be selected and the other GEBV(s) will be replaced by the corresponding EBV(s). An S-GMACE pilot run has been performed at the Interbull Centre in Uppsala. The goal of this study is to describe the data and GEBV-selection procedure and to present the results of the pilot runs.

Materials and Methods

Countries having passed the official *GEBV test* (Nilforooshan *et al.,* 2010) were asked to send GEBV data for Holstein bulls on five traits from different trait groups to Interbull Centre in the file format 01X (Interbull, 2009) until January 21, 2011. The traits were protein yield (**pro**), somatic cell score (**scs**), stature (**sta**), direct longevity (**dlo**) and cow conception 1 (**cc1**) measured as conception rate (Jorjani, 2007a,b).

Additional explanatory codes were introduced: young bulls with parent average and genomic information only should be coded as type of proof = 13; and genomically selected young bulls should be coded as status = 15. To distinguish the genomic records from the conventional ones, the first byte of GEBV data

files should be '7' instead of '0'. The genomic data corresponded to the conventional data sent in for the December 2010 routine evaluation with the inclusion of young bulls. The pedigree extraction was from February 10, 2011 and included such young bulls.

The Interbull data verification and data preparation programs (Jakobsen and Hjerpe, 2006) were adapted to the genomic requirements which were: a) no restriction on number of herds and daughters; b) no restriction on status of bull and type of proof; c) accommodation of new categories for status and type of proof for young bulls.

GEBV data for 1 to 5 traits were received from 7 countries ranging from about 48,000 for cc1 (3 countries) to 76,000 for pro (7 countries). The proportion of young bulls in these data was about 30-40% (Table 1).

1. Data preparation

The incoming data were checked for consistency. In the case where GEDCs were larger than the corresponding EDCs while genomic reliabilities were larger than the conventional ones, the genomically added EDC was estimated from the added genomic reliability (genomic reliability – conventional reliability) and added to the conventional EDC (GEDC = EDC + added genomic EDC).

2. Edits

GEDCs were estimated from genomic reliability

$$GEDC=(4-h^2)/h^2 * (GREL/(1-GREL)$$

where h^2= heritability estimate of the respective trait in the respective country; GREL=genomic reliability.

Therefore, bulls with a genomic reliability of 99% had a lower GEDC than EDC due to the rounding limit of 99% reliability. In these cases the EDC instead of the GEDC was used. GEBVs were discarded when GEDC was lower than the corresponding EDC and reliability was lower than 99%. GEBVs were excluded when the national publication of a GEBV was not official while the publication of the corresponding EBV was official.

The corresponding conventional data were pre-edited according to the criteria in the conventional MACE (**CMACE**) runs (breed specific threshold for number of herds and daughters, import bulls; Interbull, 2008).

1. GEBV selection

In the S-GMACE procedure only one GEBV per bull may enter the analysis. Therefore, in the case that more than one country submits a GEBV for the same bull only one GEBV must be selected.

GEBV selection should be based on the accuracy of the estimate. Theoretically GEDC or genomic reliability would be the criteria of choice. For within country evaluation these might be the most appropriate criteria. However, levels of GEDC and genomic reliability for bulls with no daughters differed across countries (Table 2). Factors affecting such differences could be, among others, different national evaluation procedures and different sizes of the reference population. The problem became most obvious when bulls provided by several EuroGenomics countries (David *et al.,* 2010), which have a common reference population, had lower GEDCs and genomic reliabilities in the country of origin where they had daughters than in another country where they did not have daughters.

Bulls with several GEBVs in the data were selected hierarchically according to the following criteria:

1. Bull has daughters in at least one country: Select GEBV from the country with the largest number of daughters.

2. Bull has no daughters and is submitted by the country of first registration: Select GEBV from the country of first registration.

3. Bull has no daughter and is not submitted by the country of first registration: Select GEBV from the country with largest correlation to the country of first registration.

4. Bull has no daughters and is not submitted by the country of first registration; country of first registration does not participate in international evaluation: Select GEBV from the country with the largest genomic reliability.

About 98% of the bulls with multiple GEBV for protein yield had daughters in at least one country (category 1). There were only few bulls or none in category 4, i.e., bulls without daughters with first registration in a country that does not (yet) participate in international evaluations (Table 3).

Genomic data with only one GEBV per bull and the corresponding pre-edited conventional data were merged. In case of both a GEBV and an EBV for the same bull within country the EBV was discarded. The number of bulls in the input data for S-GMACE is shown in Table 4. The proportion of genomic data varied according to country and trait. The final dataset included 68,481 (cow conception 1) to 151,904 records (protein yield) with an overall GEBV proportion from 40% (protein yield) to 52% (cow conception 1).

3. Pilot runs

For the pilot runs MACE methodology (Schaeffer, 1994) using sire-dam relationships had been applied. In total four runs were performed: two runs including the genomic data and 2 runs with the corresponding conventional data: 1) S-GMACE run including only the countries providing GEBVs for the respective trait (**gr**); 2): CMACE run including the same countries as the gr-run but only conventional EBVs (**cr**); 3) S-GMACE run including all the countries participating in the MACE evaluation in December 2010; this was the same as the gr-run but extended to include also countries not providing GEBVs for the pilot (**ga**); 4) CMACE run including conventional data from all the countries participating in the MACE evaluation in December 2010 for the respective traits (**ca**).

Results

While all the bulls submitted to the different runs were considered, only bulls with proofs coded as official (Y) were distributed.
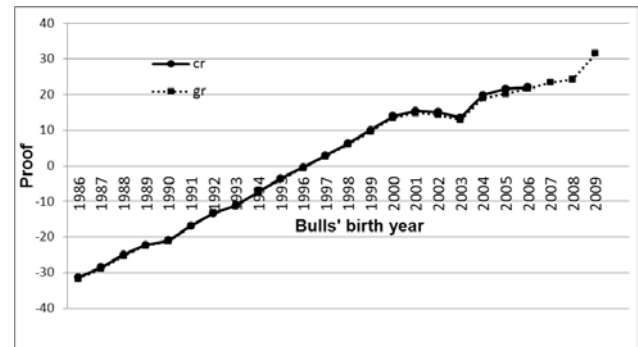

Figure 1. Genetic trend for protein yield on DEU scale for reduced S-GMACE (gr) and CMACE (cr).

The genetic trends for protein yield on the DEU scale for both S-GMACE run (gr) and CMACE run (cr) were almost identical until 2006 where the trend of cr was truncated. The trend of gr after 2006 continued smoothly until 2008, while the trend for 2009 seemed to increase (Figure 1). The same was true for the runs ga and ca where all countries participating in international evaluations were included (results not shown).

With the inclusion of GEBVs there was an overall increase of reliability for all the populations participating in the international evaluation. The mean increase in the reduced data set ranged from 0.6% for pro to 2.1% for dlo. Maximum average increase for pro and dlo was on the USA scale, lowest increase on the scale of New Zealand. The increase was mainly due to the genomic contribution of younger bulls with few daughters born in 2005 and 2006 (Figure 2).

Differences in genomic reliabilities among countries were also visible in the S-GMACE output data (Table 5): Genomic reliabilities on the USA scale were larger than on the other country scales. Within EuroGenomics countries differences were noticeable for cc1.

3

Among the top 100 bulls for protein yield the number of young bulls ranged from 24 (Polish scale) to 64 (scale of New Zealand). Most of those young bulls had their first registration in the USA on all country scales. This was not surprising as the USA provided about 75% of all the official young bulls. Also young bulls with first registration in non-GEBV providing countries were among the top 100. Average international reliabilities of the young bulls on the top-100 list ranged from 63 (scales of Poland and New Zealand) to 76 on the USA scale (Table 6). Correlations among countries, GEDCs, size of reference population as well as national genomic evaluation model are factors that affect these values.

Correlations between national and international proofs were generally larger than 0.99 across traits and country scales (Table 7). The respective correlations between national and international reliabilities were a bit lower and more spread showing differences both among country scales and traits. Lowest correlations among national and international reliabilities were on the scale of Poland ranging from 0.778 (sta) to 0.868 (pro); largest correlations among national and international reliabilities were on the USA scale with a range from 0.970 for scs to 0.980 for pro.

Summary

There seemed to be differences in the estimation procedures of genomic reliability across countries. Therefore other criteria than GEDC or genomic reliability, i.e., number of daughters or country of first registration were used for GEBV selection. Harmonization of input data is important in order to fairly select among GEBVs from different sources.

Results of S-GMACE were conclusive. Genetic trends of breeding values of conventional and genomic runs were almost identical. The trend for young bulls (genomic data only) was following the same trend line. Young bulls with no daughters appeared on the top-100 lists on several country scales.

There was an overall increase in international reliabilities with S-GMACE on all country scales including also non-GEBV providing countries.

Acknowledgment

References

David, X., de Vries, A., Feddersen, E. & Borchersen, S. 2010. International Genomic Cooperation. EuroGenomics significantly improves reliability of Genomic evaluation. *Interbull Bulletin 41,* 77-78.

Interbull. 2008. *Interbull Code of Practice.* Section 5. Method of international evaluation. http://www.interbull.org/images/stories/chapter_5.pdf

Interbull. 2009. *Service Documentation.* General forms and Documents. http://www.interbull.org/index.php?option=com_content&view=article&id=55&Itemid=76

Interbull. 2010. *Interbull Routine Genetic Evaluation for Female Fertility Traits.* December 2010. http://www-interbull.slu.se/Female_fert/framesida-fert.htm

Jakobsen, J.H. & Hjerpe, E. 2006. Interbull checks of incoming data. *Interbull Bulletin 34,* 11-15.

Jorjani, H. 2007a. International Genetic Evaluation of Female Fertility Traits in Five Major Breeds. *Interbull Bulletin 37,* 144-147.

Jorjani, H. 2007b. There and Back Again: A Tale of Choosing Female Fertility Traits. *Interbull Bulletin 37,* 148-151.

Nilforooshan, M., Zumbach, B., Jakobsen, J., Loberg, A., Jorjani, H. & Dürr, J. 2010. Validation of national genomic evaluations. *Interbull Bulletin 42,* 56-61. http://www.interbull.org/images/stories/G_Validation_HJ.pdf

Schaeffer, L.R. 1994. Multiple-Country Comparison of Dairy Sires. *J. Dairy Sci. 77,* 2671-2678.

Sullivan, P.G. & VanRaden, P.M. 2010. GMACE implementation. *Interbull Bulletin 41,* 3-7.

VanRaden, P.M. & Sullivan, P.G. 2010. International genomic evaluation methods for dairy cattle. *Genetics Selection Evolution 42,* 7.

Table 1. GEBV data by country and trait[1] received at Interbull Centre.

Country[2]	pro		scs		sta		dlo		cc1	
	all[3]	YB[4]	all	YB	all	YB	all	YB	all	YB
DEU	28,645	9,346	28,645	9,346	28,645	9,346	28,645	9,346	28,645	9,346
DFS	5,962	1,178	-	-	-	-	-	-	-	-
FRA	10,274	4,745	14,704	4,762	14,704	4,624	14,704	4,804	14,704	4,991
NLD	5,026	866	5,028	866	5,028	847	5,028	753	5,028	870
NZL	2,896	349	2,896	349	2,896	349	-	-	-	
POL	2,237	61	2,543	367	2,298	262	-	-	-	
USA	21,069	11,346	21,069	11,346	-	-	21,069	11,346	-	-
Total	76,109	27,891	74,885	27,036	53,571	15,428	69,446	26,249	48,377	15,207

[1]pro=protein yield; scs=somatic cell score; sta=stature; dlo=direct longevity; cc1=cow conception 1.
[2]DEU= Germany; DFS=Denmark, Finland and Sweden; FRA=France; NLD= The Netherlands; NZL= New Zealand; POL= Poland; USA= The United States of America.
[3]all= all bulls; [4]YB= young bulls (type of proof=13).

Table 2. Mean GEDC and genomic reliability (g_rel) of proofs of young bulls (type of proof = 13) for different traits[1].

Country[2]	pro		scs		sta		dlo		cc1	
	GEDC	g_rel	GEDC	g_rel	GEDC	g_rel	GEDC	g_rel	GEDC	g_rel
DEU	17	73	46	77	17	71	22	53	112	44
DFS	12	63	-	-	-	-	-	-	-	-
FRA	30	70	54	68	16	69	42	54	301	61
NLD	12	63	17	64	15	71	26	45	69	44
NZL	20	62	37	59	15	60	-	-	-	-
POL	13	60	13	60	5	59	-	-	-	-
USA	57	80	109	75	-	-	115	69	-	-

[1]pro=protein yield; scs=somatic cell score; sta=stature; dlo=direct longevity; cc1=cow conception 1.
[2]DEU= Germany; DFS=Denmark, Finland and Sweden; FRA=France; NLD= The Netherlands; NZL= New Zealand; POL= Poland; USA= The United States of America.

Table 3. Number of bulls and records by GEBV selection categories for protein yield.

Categories for GEBV selection	Number of bulls		Number of records[1]
	unique	multiple	multiple bulls
1. Bulls with daughters	20,612	12,851	26,631
2. Bulls without daughters from country of 1st registration	19,580	247	506
3. Bulls without daughter from country with largest r_g[2] to country of first registration	8,561	67	139
4. Bulls without daughters with first registration in a country not participating in international evaluations	1,670	0	0

[1]Number of records from unique bulls = number of unique bulls; [2] genetic correlation.

Table 4. Number of EBVs and GEBVs by country in the S-GMACE input data for different traits[1].

Country[2]	pro		scs		sta		dlo		cc1	
	EBV	GEBV	EBV	GEBV	EBV	GEBV	EBV	GEBV	EBV	GEBV
DEU	16,922	16,001	16,863	19,732	13,800	20,294	15,740	20,288	14,872	20,549
DFS	6,751	5,051	-	-	-	-	-	-	-	-
FRA	15,365	9,955	10,614	10,794	12,171	10,680	13,451	10,249	8,266	10,485
NLD	10,283	4,961	9,815	4,967	9,157	4,969	8,954	4,988	9,310	4,999
NZL	4,687	2,756	3,829	2,748	2,625	2,795	-	-	-	
POL	4,626	2,045	3,281	2,317	3,423	2,151	-	-	-	
USA	31,352	20,699	24,420	20,706	-	-	34,646	20,733	-	-
Total	89,986	61,918	68,822	61,264	41,176	40,889	72,791	56,258	32,448	36,033

[1]pro=protein yield; scs=somatic cell score; sta=stature; dlo=direct longevity; cc1=cow conception 1.
[2]DEU= Germany; DFS=Denmark, Finland and Sweden; FRA=France; NLD= The Netherlands; NZL= New Zealand; POL= Poland; USA= The United States of America.

Table 5. Average international reliabilities of young bulls (type of proof = 13) with official proofs for the respective national data[1] by trait[2] and country on national country scale (distributed).

Country[3]	pro	scs	sta	dlo	cc1
DEU	75	77	72	60	47
FRA	74	71	73	60	65
NLD	-	75	72	-	56
NZL	68	65	67	-	-
USA	82	78	-	73	-

[1]Only data provided by the respective country.
[2]pro=protein yield; scs=somatic cell score; sta=stature; dlo=direct longevity; cc1=cow conception 1.
[3]DEU= Germany; FRA=France; NLD= The Netherlands; NZL= New Zealand; USA= The United States of America.

Table 6. Number of bulls born in 2007 or later in the top 100 for protein yield on different country scales and mean reliability (rel).

Country scale[1]	N	Country of first registration												rel
		LUX	AUS	DEU	DFS	FRA	HUN	NZL	NLD	CAN	USA	ITA	GBR	
DEU	29		1	8	2	2	1		3		12			70
DFS	29	1		6		3			5		13		1	67
FRA	40		2	4	1	13			4	1	14	1		68
NLD	25			4		3			4		12		2	66
USA	52	1		4	1	4			7	2	31		2	76
NZL	64		1	2		1		50	3		6		1	63
POL	24		1	3		3			4		12		1	63

[1]DEU= Germany; DFS=Denmark, Finland and Sweden; FRA=France; NLD= The Netherlands; NZL= New Zealand; POL= Poland; USA= The United States of America.

Table 7. Correlations between national and international proofs (r_proof) and reliabilities (r_rel) in S-GMACE for different traits[1] and different countries[2] - results from the reduced run (gr).

Country[2]	pro		scs		sta		dlo		cc1	
	r_proof	r_rel	r_proof	r_rel	r_proof	r_rel	r_proof	r_rel	r_proof	r_rel
DEU	0.999	0.948	0.996	0.883	0.999	0.944	0.997	0.909	0.987	0.809
DFS	0.998	0.815	-	-	-	-	-	-	-	
FRA	1.000	0.915	0.998	0.933	0.999	0.931	0.993	0.941	0.999	0.978
NLD	0.999	0.893	0.994	0.822	0.999	0.916	0.988	0.926	0.998	0.913
NZL	0.997	0.955	0.988	0.846	0.999	0.930	-	-	-	-
POL	0.999	0.868	0.990	0.814	0.998	0.778	-	-	-	-
USA	0.999	0.980	0.997	0.970	-	-	0.996	0.977	-	-

[1]pro=protein yield; scs=somatic cell score; sta=stature; dlo=direct longevity; cc1=cow conception 1.
[2]DEU= Germany; DFS=Denmark, Finland and Sweden; FRA=France; NLD= The Netherlands; NZL= New Zealand; POL= Poland; USA= The United States of America.
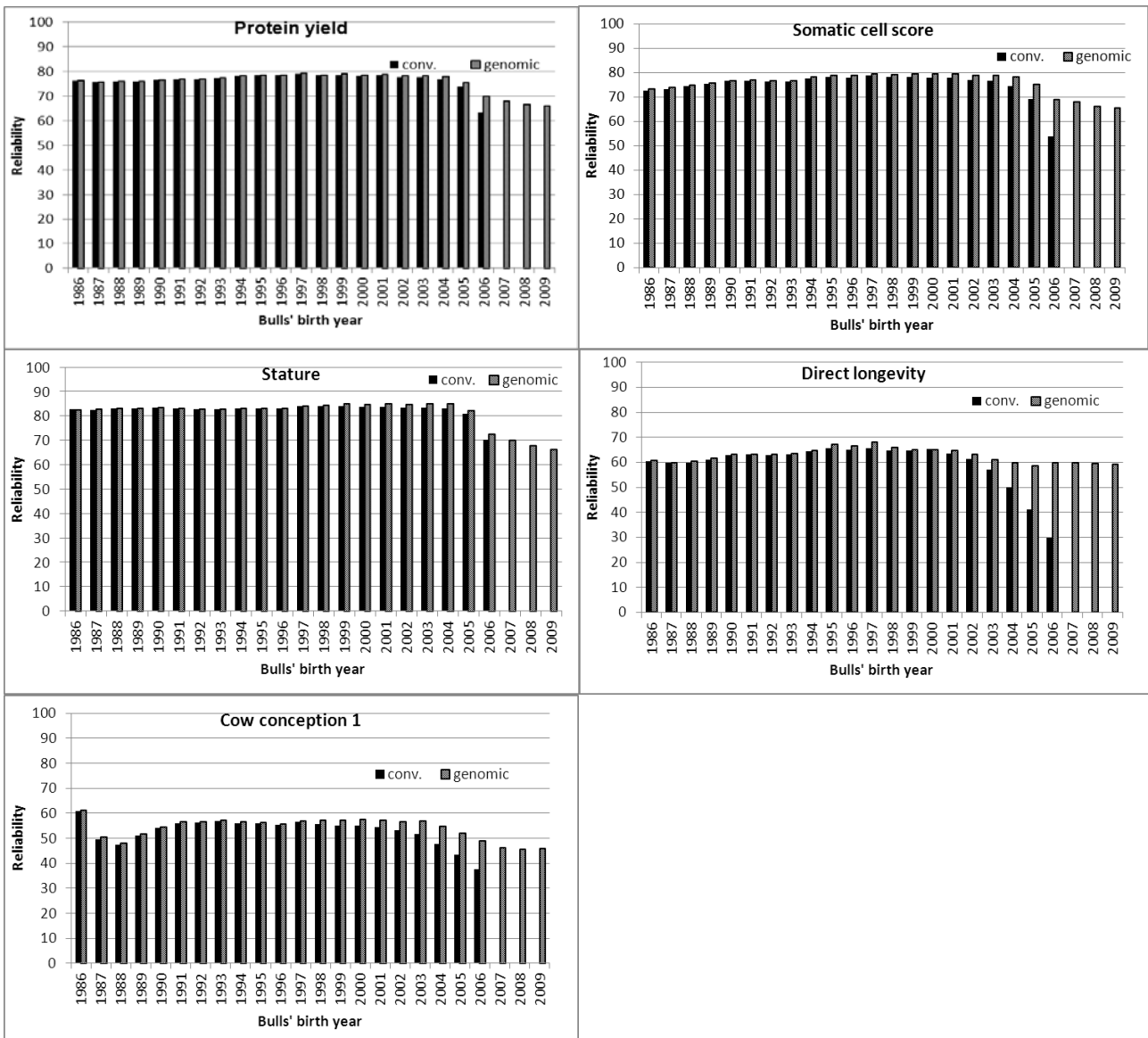
Figure 2. Average international reliability per birth year of bull for the conventional reduced (conv) and the genomic reduced (genomic) runs for protein yield, somatic cell score, stature, direct longevity and cow conception 1 on German scale.