Validation and Implementation of New Genomic Traits in The Netherlands: Lactose, Urea, Calf Survival, Ketosis, Conception Rate, Heifer Fertility, and AMS Traits

W.M. Stoop, C. Schrooten, I. Vermeer and G. de Jong
CRV, P.O. Box 454, 6800 AL Arnhem, The Netherlands
E-mail: Marianne.Stoop@crv4all.com

Abstract

Since April 2015, the Dutch-Flemish evaluation has been publishing genomically enhanced breeding values (GEBV) for a total of 65 traits and indices. These include traditional traits like milk production, conformation, longevity and calving traits; health traits like udder health, claw health, and fertility traits; as well as novel traits that were recently introduced in the Dutch-Flemish (genomic) evaluation: lactose, urea, conception rate (to replace non-return 56), heifer fertility traits, calf survival, ketosis, and automatic milking system (AMS) traits.

In the Dutch-Flemish genomic evaluation systems bulls are considered for the training population when their breeding value for the trait of interest exceeds 50 percent reliability. However, additional edits might be advisable when correlations between countries are low or for traits where the reliability is highly dependent on correlated traits, rather than true (daughter) observations.

Key words: genomics, validation, dairy cattle, new traits

1. Introduction

Starting April 2015, the Dutch-Flemish evaluation has been publishing genomically enhanced breeding values (GEBV) for a total of 65 traits and indices. These include traditional traits like milk production, conformation, longevity and calving traits; health traits like udder health and claw health, and fertility traits; as well as novel traits that were recently introduced in the Dutch-Flemish (genomic) evaluation: lactose, conception rate (to replace non-return 56), heifer fertility traits, calf survival, ketosis, and automatic milking system (AMS) traits. For some traits validation results tend to be lower than expected based on the heritability of the trait and the number of bulls in the training population. There are several possible causes. Firstly, validation bulls might be less connected to the training population. Secondly, phenotypes of training population bulls might be less accurate. This might be the case if phenotypes (de-regressed conventional breeding values) are considered of both national and international bulls, while the correlation between foreign countries and The Netherlands is below 0.8; or when the conventional breeding value (EBV) was estimated in a multi-trait model where the actual EBV is highly dependent on correlated traits, rather than daughter observations on the trait itself. This paper describes the validation of nine new traits implemented in the Dutch-Flemish evaluation, as well as two alternative validations (one for lactose yield and one for locomotion), deleting that part of the data that is considered to be possibly flawed due to low country correlations (locomotion) or derived phenotypes (lactose yield).

2. Material and Methods

2.1 Validation procedure

For the genomic (validation) evaluation, data of the Eurogenomics consortium is used, consisting of almost 30,000 bulls in the training population for production traits. Bulls are included in the training population for a trait when their conventional (national or MACE) daughter breeding value has a reliability of at least 50 percent.

All bulls born after 20040630 are considered for the validation population and are deleted from the training population.

Direct Genomic Values (DGV) are calculated in a genomic evaluation following the method described by De Roos *et al.* (2009), using Bayes Stochastic Search Variable Selection (Bayes-SSVS) to estimate the SNP effects.

For validation bulls the conventional daughter breeding value is not included in the training population. In the validation procedure, their estimated Male Pedigree Index (PI) and DGV are both correlated to their realized daughter EBV, giving a regression coefficient r_{pi} and r_{dgv}. The squared correlations (R^2_{pi} and R^2_{dgv}) give a measure for the corresponding reliabilities. Using the EDC equivalents of these reliabilities an added reliability can be calculated for the DGV, which is then adjusted for the mean reliability of the validation cohort R^2_{val} and extrapolated to the size of the training population during routine evaluation.:

$R^2_{dgv} => EDC_{dgv}$, and $R^2_{pi} => EDC_{pi}$

$EDC_{dif} = EDC_{dgv} - EDC_{pi}$

$EDC_{dif} => R^2_{dif}$

$R^2_{cor} = R^2_{dif}/R^2_{val}$

$R^2_{add} = N_2/(N_2+N_1*(1- R^2_{cor})/ R^2_{cor})$

where N_1 is the size of the training population during validation, and N_2 is the size of the training population during routine evaluation.

2.2 Trait definitions

Trait definition of the new traits:
Lactose yield is expressed in kilograms. For bulls without lactose observations (international bulls), but with milk yield present, lactose yield is derived as milk yield times average lactose percentage. Urea is measured in mg per 100 ml milk. Conception rate for cows is send to Interbull as T3 (replacing non return 56) and conception rate for heifers is send to Interbull as T1. Both are calculated as 100/#inseminations per pregnancy, with observations missing if gestation length is less than 54 days and observation is 0 when gestation length is over 300 days (failed conception). Age at first insemination is measured in days. Calf survival is defined as the first year survival of replacement heifers in the dairy sector and is measured as the percentage survival from day 3-365. Ketosis is estimated based on a formula using acetone, mBHBA, and fat and protein corrected fat/protein ratio, according to Van der Drift (2012). For automatic milking systems (AMS), three traits are available in the conventional national evaluation: AMS efficiency, milking interval, and habituation of heifers. However, the last trait lacks the minimum amount of data for a proper genomic evaluation and is therefore not included. AMS efficiency is defined as the total milk yield in kilograms per total AMS time in minutes. The milking interval is defined as the number of minutes between two consecutive successful milkings.

2.3 Alternative validation

For lactose yield and locomotion an alternative validation is tested. For the alternative validation of lactose yield only bulls are considered in the training population that have daughter observations for lactose yield. This implies only (part of the) national data is included, and all foreign bulls are excluded. For the alternative validation of locomotion only national data is considered, whereas international data is deleted. The country correlations for locomotion average 0.68 between The Netherlands and foreign countries, which illustrates the difference in trait definition between countries.

3. Results & Discussion

Table 1 shows the added genomic reliability, as well as the reliability of the Genomically Enhanced Breeding Value (GEBV) of young bulls without daughters for the total merit index NVI, most underlying traits in the NVI, and claw health. The NVI reliability for young bulls without daughters is on average .64 for the final published GEBV. Largest reliabilities are seen for overall conformation (average of .70), production traits (protein averages at .67), and fertility index (.67). Added reliability from genomic information is largest for fertility index (.66). Added reliability is lowest for claw health (.32), which is not surprising, as this trait is completely based on national data, implying that the training population for claw health is much lower (approximately 3,500 bulls) than for traits including international data (approximately 30,000 bulls).

Table 1. Added reliability(%) from genomic information and total GEBV reliability of young bulls without daughters for NVI, underlying traits, and claw health.

	Add rel	GEBV rel
Overall index NVI	60	64
Production (kg prot)	60	67
Longevity	41	47
Overall conformation	49	70
Fertility index	66	67
Calving index	55	62
Udder health index	60	65
Claw health index	32	47

Table 2 shows the added genomic reliability, as well as the reliability of the Genomically Enhanced Breeding Value (GEBV) of young bulls without daughters for the nine new traits, introduced in April 2015, in the Dutch-Flemish evaluation. Of the new traits, only Lactose yield and the two conception rate traits are based on both national and international data. The other traits are based on national data only, which is limited in the number of bulls, especially for automatic milking system traits and ketosis, for which data collection has started only in recent years.

Table 3 shows the results of the standard validation procedure and the alternative validation procedure for lactose yield and locomotion. For lactose yield, results improve when only bulls with true daughter observations (and at least .50 reliability) are included in the training population. Even though in the alternative validation only approximately 10% of the bulls from the original training population set are included, added reliability is increased by 4 %. Therefore, observations derived from milk yield do not seem to add any extra information to the system. For locomotion, the alternative validation shows a dramatic loss in added reliability, from .48 to .32. Also in this alternative validation, a huge number of training bulls is lost, compared to the standard validation set. However, for locomotion, foreign bulls do add new information to the system. The low country correlation may diminish the value of the information, but this is correctly reflected in the given reliability of the foreign bulls, giving this new foreign information the correct weight in the standard validation. Deleting this foreign information diminishes the information content in the training population, resulting in much lower added reliabilities for locomotion in the alternative validation procedure.

Table 2. Added reliability (%) from genomic information and total GEBV reliability of young bulls without daughters for new traits in the Dutch-Flemish evaluation.

	# bulls in training pop.	h^2	added rel	GEBV rel
Kg Lactose	28,686	0.55	35	47
Urea	5,446	0.60	51	57
Conception rate cows	15,068	0.036	55	57
Conception rate heifers	11,422	0.018	26	41
Age at first insemination	5,776	0.045	21	36
Calf survival	5,674	0.011	28	42
AMS efficiency	3,226	0.27	35	47
AMS interval	3,296	0.12	18	35
Ketosis	3,964	0.24	52	58

Table 3. Added reliability (%) from genomic information for lactose yield and locomotion, based on standard validation procedure and the alternative validation procedure in which only national data is considered.

	#bulls in training pop.	Add rel. standard val.	#bulls in training pop.	Add rel. alternative val.
Kg Lactose	28,686	35	2,900	39
Locomotion	29,093	48	4,496	32

C**onclusion**

The Dutch-Flemish evaluation currently consists of 65 genomic traits. Typically, GEBV reliabilities of young bulls vary between .40 for individual functional traits and .70 for indices and highly heritable traits. When setting edits for bulls to enter the training population, there is a balance between information content of data and number of bulls in the training population. However, as the results have shown, adding foreign information from low correlated countries can still add information to the genomic system, resulting in increased reliabilities. Mimicking data from a highly correlated trait as milk yield (correlation between milk yield and lactose yield is .96) might give a good indication of the breeding value for lactose yield, but does not seem to add new information to the system. Deleting this data even increases the genomic predictability.

4. References

De Roos, A.P.W., Schrooten, C., Mullaart, E., van der Beek, S., de Jong, G. & Voskamp, W. 2009, Genomic Selection at CRV, *Interbull Bulletin 39,* 47-50.

Van der Drift, S.G.A., Van Hulzen, K.J.E., Teweldemedhn, T.G., Jorritsma, R., Nielen, M. & Heuven, H. 2012, Genetic and nongenetic variation in plasma and milk ß-hydroxybutyrate and milk acetone concentrations of early-lactation dairy cows. *J. Dairy Sci. 95,* 6781-6787.