

Estimating Genetic Correlations of Traits Relevant for Calculating a Total Merit Index

Z. Liu¹, J. Tarres^{1,2}, V. Ducrocq², F. Reinhardt¹ and R. Reents¹

¹vit w.V., Heideweg 1, 27283 Verden, Germany

²UR337, SGQA, INRA, 21 agronomique, 78352 Jouy-en-Josas Cedex, France

Abstract

For calculating a total merit index genetic correlations among various traits were needed. A multi-trait animal model was applied to daughter yield deviations (DYD) and their associated effective daughter contribution of bulls for estimating genetic correlations among the trait groups: milk, fat, and protein yields, somatic cell scores (SCS), longevity expressed as relative risk, female fertility traits, five linear type traits plus locomotion, and body condition scores. Genetic (co)variances of the selected traits were estimated using an approximate REML method. A total of 5709 Holstein bulls with DYD were chosen and 7903 ancestors were traced back for parameter estimation. Both genetic correlations and variances converged well. Moderate to high genetic correlations were found among the three production traits. Genetic correlation between production trait and SCS changed from positive to negative during the first lactation and from nearly zero to negative in the later lactations. While non-return rates (NR) were almost uncorrelated with the production traits, the interval fertility traits, interval calving to first insemination (CF) and interval first to successful insemination (FS) were moderately correlated with the yield traits. Low genetic correlations were found between production and the type traits, except udder depth. Functional longevity showed very low genetic correlations with the yield traits. SCS were proven to be a good predictor of longevity with a genetic correlation of 0.59. The interval fertility traits were moderately correlated with functional longevity and can be regarded as a second best predictor of longevity. The genetic correlation estimates will be used in a 2-step multi-trait animal model for deriving various selection indices, such as total merit index.

1. Introduction

Total merit index (TMI), a function of estimated breeding values (EBV) of economically important traits, are widely used in dairy cattle breeding programmes worldwide. TMI typically contains information from several component traits: milk production traits, conformation traits including locomotion, health traits like SCS und body condition score (BCS), longevity, calving traits, workability, and female fertility. Usually countries evaluate the trait groups separately using different statistical models, such as a random regression test day model (RRTDM) for production or SCS traits (Liu *et al.*, 2004), a non-linear survival model for longevity (Ducrocq 2001, Tarres *et al.*, 2006), or a multiple trait model for female fertility (Liu *et al.*, 2008). Most countries, except France, apply so far the standard selection index method to combine EBV of these component traits from single trait models. A 2-step animal

model (AM2) using pre-corrected records (Ducrocq *et al.*, 2001) offers the following advantages over the selection index approach for setting up TMI. Indirect selection effect on correlated traits can be optimally considered. Higher genetic progress and proof reliabilities have been confirmed for the AM2 model (Lassen *et al.*, 2007). Double counting of some trait information is avoided, e.g. SCS proof included in combined longevity prediction as well as in TMI that contains again the same SCS information and combined longevity. All cows will also receive longevity proofs, whereas survival analysis model predicts longevity proofs only for bulls. Compared to yield deviations of cows, DYD of bulls allow a more efficient estimation and more reliable estimates of genetic correlations among traits as a result of their much higher reliability. The objective of this study was to estimate genetic correlations among the component traits of the German TMI using DYD and associated effective daughter contribution (EDC) of bulls.

2. Materials and Methods

2.1. Data materials

Data from April 2008 routine genetic evaluations for production, SCS, conformation, longevity and female fertility traits were used for estimating genetic correlations among the traits. Bulls' DYD of those traits were calculated following the multi-trait model procedure (Liu *et al.*, 2004), in addition to their associated EDC matrices or scalars, depending on genetic evaluation models of the analysed traits, e.g. a vector of DYD and a matrix of EDC for female fertility traits of bull (Liu *et al.*, 2008). Because the three production and SCS traits were evaluated with a RRTDM, DYD of bulls were expressed as random regression coefficients (RRC) of Legendre polynomials with three parameters. For longevity evaluated with a survival model, a pseudo-record of relative risk and its weight were calculated for each daughter of a bull following a procedure by Ducrocq (2001) and Tarres *et al.* (2006), which were then used to compute DYD and EDC for the bull. Calving and workability traits were not considered in this study. The bulls were required to be present in Interbull's 010 file for production traits, 015 file for conformation, 016 file for SCS, 017 file for longevity, and 019 file for female fertility. However, due to much shorter history of data recording for locomotion and BCS, no restriction was imposed on availability of these two traits. In vit's multi-breed national models for longevity and female fertility, breed effect was not considered, and consequently the breed difference in these traits was contained in their DYD. In contrast, DYD of production, SCS and conformation traits were free of the breed differences. Because the software for estimating the genetic correlations (Tarres *et al.*, 2007) does not permit fitting a second fixed effect, e.g. breed effect, in its present version, only Black-and-White Holstein bulls were chosen for the parameter estimation to avoid fitting of the additional breed effect. Besides the selection on presence of the traits and breed of bulls, further data editing dealt with lactation length of daughters in production and SCS traits. As the multi-lactation RRTDM (Liu *et al.*, 2004) provided DYD in first three lactations, adequate daughter information was needed for

estimating genetic correlations between each of the lactations with the other traits. Therefore, bulls were required to have daughters' test day records in all three lactations. Additionally, all bulls must have no fewer than 30 daughters with lactation passing 120 days in milk in each of the three lactations. No further selection was imposed on the remaining traits evaluated other than the test day model. A total of 5709 bulls with DYD remained after all the selection steps. Table 1 describes the final data set for the parameter estimation. The number of the operational traits reached 49, with 9 RRC for each production or SCS trait.

Bull pedigree file from Interbull's April 2008 evaluations was reformatted from a sire, maternal grand-sire and maternal grand-dam format to a sire and dam format. Ancestors of the selected bulls with data were traced back from both sire and dam sides as far as possible. The final pedigree file contained 13,612 animals plus 18 phantom parent groups which were formed according to breed, country of origin, selection paths (son to sire, son to dam, daughter to sire and daughter to dam) and birth year of the animal. Small phantom groups were merged to ensure at least 200 animals assigned to each group. Among the 7903 ancestors, 281 sires had also DYD data available.

2.2. Statistical model

The following statistical model was applied to estimate genetic correlations among the selected traits:

$$\mathbf{q}_{ij} = \mathbf{f}_{jk} + \mathbf{a}_{ij} + \mathbf{e}_{ij} \quad [1]$$

where \mathbf{q}_{ij} is a vector of DYD of the i -th bull in trait j , \mathbf{f}_{jk} is a vector of fixed effects of birth year k in the j -th trait, \mathbf{a}_{ij} is a vector of additive genetic effects of bull i in trait j , and \mathbf{e}_{ij} is a vector of residual effects. Adding a birth year effect in the model can provide more robust estimation of genetic trends (Lassen *et al.*, 2007). For traits evaluated with a single trait model, e.g. longevity, all the vectors become scalar. The (co)variance matrix of genetic effects of the m component trait blocks is denoted as:

$$\mathbf{G}_0 = \{\mathbf{G}_{0jl}\}_{j=1,\dots,m;l=1,\dots,m} \quad [2]$$

where $\mathbf{G}_{0_{jj}}$ is genetic (co)variance matrix of trait j , and $\mathbf{G}_{0_{jl}}$ is the genetic covariance matrix between traits j and l . The inverse of error (co)variance matrix of bull i in trait j is:

$$[\text{Var}(\mathbf{e}_{ij})]^{-1} = \Psi_{ij} \quad [3]$$

where Ψ_{ij} is EDC matrix for bull i in trait j on animal basis, converted from reliability matrix contributed by his daughters' records in the j -th trait. The multi-trait EDC method (Liu *et al.*, 2004) was applied to approximate matrix Ψ

for all the bulls. Similar to the parameter estimation for a multi-trait MACE model (Tarres *et al.*, 2007), residual correlations between DYD of two traits were ignored, because it can be verified that the proportion of residual covariance in the covariance of DYD between two traits decreases with the number of daughters of the DYD. Even for two traits with high residual correlation, the residual covariance between the two DYD will become negligible when the number of daughters of the DYD is greater than 100. Therefore, the residual correlation of the DYD was not considered in the estimation of genetic (co)variances.

Table 1. Number of Black-and-White Holstein bulls and average number of daughters by birth year for parameter estimation.

Year of birth	No. of bulls	Average no. of daughters in					Average no. of daughters in		
		Milk, fat protein	SCS	Type traits	Longevity	Female fertility	No. of bulls	Loco-motion	BCS
1986	63	6219	6214	701	4429	2606	2	21	22
1987	59	2918	2916	299	2124	1530			
1988	93	2339	2340	244	1668	1322	2	32	34
1989	110	3918	3918	523	2931	2471	9	455	545
1990	148	3566	3566	543	2571	2262	15	287	320
1991	186	2715	2716	408	2026	1822	32	66	75
1992	244	1405	1405	253	1044	958	24	204	226
1993	639	570	570	107	426	386	41	169	199
1994	694	728	728	132	527	510	70	302	350
1995	716	211	211	66	154	144	47	147	181
1996	648	125	125	51	96	86	12	47	54
1997	661	115	115	52	88	79	6	20	21
1998	619	115	115	54	88	80	3	40	41
1999	575	123	123	59	95	87	260	30	36
2000	252	122	122	59	95	88	252	45	53
2001	2	99	99	67	84	67	2	52	66
All	5709	690	690	126	505	555	777	91	107

Mixed model equations of model 1 were solved using a pre-conditioned conjugate gradient algorithm and an iteration on data technique. An approximate expectation maximization REML method was implemented to estimate the across-trait genetic correlations (Tarres *et al.*, 2007). The iterative process of the parameter estimation was considered converged when the third decimal place of all the genetic correlation estimates no longer changed between two consecutive rounds of iteration.

3. Results

The parameter estimation was run on a 64-bit AMD Opteron Linux server. The REML program took about 700 Mb RAM and 32 CPU minutes per iteration round. A total of 349 rounds were needed to reach the pre-defined convergence criterion. The (co)variance estimates on RRC basis of production traits and SCS were converted to 305-day single lactations as well as combined lactation (Liu *et al.*, 2004). Genetic correlations of days open

(DO) were derived from its two component traits CF and FS.

Figure 1 shows estimated genetic correlations between milk yield and SCS in first three lactations. As found in other studies (Haile-Mariam *et al.*, 2001, Reents *et al.*, 1994), the genetic correlations changed from positive to negative during the first lactation and from nearly zero to negative in second and third lactations. The genetic correlations between milk yield and SCS were not consistently unfavourable. In contrast to the correlation estimates between milk yield and SCS between lactations, estimated genetic correlations between milk and fat yields, shown in Figure 2, were more similar between the lactations. The two yield traits were less positively correlated in the middle of lactation than at beginning or at end of the lactation. Milk and protein yields were significantly higher correlated than milk and fat yields and also higher than fat and protein yields. SCS had similar genetic correlation patterns with fat or protein yields as milk yield. From the figures we can see that using DYD in form of RRC can model the complex genetic correlations among the traits more appropriately than using DYD on a lactation basis.

3.1. Genetic correlation estimates of production traits

Among the three production traits, high genetic correlations can be seen in Table 2, with the exception of genetic correlation between milk and fat yields, the lower correlation was caused by low genetic correlations between milk and fat yield from different lactations. On the combined lactation level, production traits were very low correlated with SCS. Table 2 shows that higher yield is associated with higher culling risk at a very low level, and genetic correlation changed from 0.26 in the first lactation to -0.06 in the third lactation (Table 3). Production traits were moderately correlated with interval fertility traits FS, CF and DO, whereas their genetic correlations with NR rates were close to 0 as shown in Table 2. The genetic correlations to cow fertility traits CF and consequently to DO decreased gradually from first to third lactation milk yield (Table 3). Overall, low genetic correlations were estimated between

production and type traits including locomotion, except that the genetic correlations with udder depth ranged from -0.18 to -0.24. The genetic correlations dropped over lactations, except the subjective score traits overall udder and feet and legs, because type classification was recorded on first lactation cows. Production traits were moderately correlated with BCS, and higher production was associated with slimmer body.

3.2. Genetic correlation estimates of SCS

With a genetic correlation estimate of 0.59 to relative culling risk, SCS were confirmed to be good predictor of longevity. The genetic correlation increased with increasing lactation number (Table 3). Although SCS were almost uncorrelated with the fertility traits NR56 and the two heifer fertility traits, genetic correlation of about 0.20 was found between SCS and interval fertility traits CF or DO. Lower genetic correlation was estimated for SCS with the cow interval fertility trait FS. No noticeable difference in genetic correlations with the other traits was found for SCS between lactations (Table 3). Genetic correlation between SCS and BCS was weak, -0.11, on a combined lactation basis.

3.3. Genetic correlation estimates of female fertility

Low genetic correlation estimates were obtained between the fertility traits and type traits, except that udder depth had a genetic correlation of about -0.22 with the fertility traits with the fertility traits CF or DO (Table 3). Among all the fertility traits, DO had the highest genetic correlation with BCS, -0.27.

3.4. Genetic correlation estimates of longevity

Based on the genetic correlation estimates with the relative culling risk, the following trait blocks, as longevity predictors, can be ranked as SCS, interval fertility traits, type traits and BCS. The three interval fertility traits had genetic correlations ranging from 0.40 to 0.58 with the relative risk. With the exception of rump angle, all the type traits including locomotion were moderately correlated with functional longevity. BCS had a relatively low

genetic correlation, -0.13, with the relative risk.

4. Discussion

Genetic correlations among a total of 49 traits were estimated by applying an approximate REML method to DYD of 5709 bulls. This estimation procedure was proven to be efficient and led to reliable parameter estimates, and genetic correlation and variance estimates converged well. Because of high number of daughters of the bulls, residual correlations among the traits could not be estimated using their DYD, which can be instead obtained with cows' yield deviations. Ignoring the residual correlations among the traits in the parameter estimation should not bias the genetic correlation estimates, because all of the selected bulls had hundreds of daughters and the proportion of residual covariance in the covariance between DYD of two traits was negligibly small. The complex genetic correlation structure between production and SCS traits demonstrated that using DYD in form of RRC can lead to more appropriate modelling than using DYD on a 305-day lactation basis.

5. References

- Ducrocq, V. 2001. A two-step procedure to get animal model solutions in Weibull survival models used for genetic evaluations on length of productive life. *Interbull Bulletin* 27, 147-152.
- Ducrocq, V., Boichard, D., Barbat, A. & Larroque, H. 2001. Implementation of an approximate multitrait BLUP to combine production traits and functional traits into a total merit index. *52nd Annual Meeting EAAP*, Budapest, Hungary, 2001.
- Haile-Mariam, M., Bowman, P.J. & Goddard, M.E. 2001. Genetic and environmental correlations between test-day somatic cell count and milk yield traits. *Lives Prod Sci.* 73, 1-13.
- Lassen, J., Sorensen, M.K., Madsen, P. & Ducrocq, V. 2007. An approximate multitrait model for genetic evaluation in dairy cattle with a robust estimation of genetic trends. *Genet. Sel. Evol.* 39, 353.
- Liu, Z., Reinhardt, F., Bünger, A. & Reents, R. 2004. Derivation and calculation of approximate reliabilities and daughter yield deviations of a random regression test-day model for genetic evaluation of dairy cattle. *J. Dairy Sci.* 87, 1896-1907.
- Liu, Z., Jaitner, J., Reinhardt, F., Pasman, E., Rensing, S. & Reents, R. 2008. Genetic evaluation of fertility traits of dairy cattle using a multiple trait animal model. *J. Dairy Sci.* (accepted)
- Reents, R., Dekkers, J.C.M. & Schaeffer, L.R. 1994. Genetic parameters of test-day somatic cell counts and production traits. *5th WCGALP*, Guelph, Canada, Vol. 17, 120-123.
- Tarres, J., Piedrafita, J. & Ducrocq, V. 2006. Validation of an approximate approach to compute genetic correlations between longevity and linear traits. *Genet. Sel. Evol.* 38, 65-85.
- Tarres, J., Liu, Z., Ducrocq, V., Reinhardt, F. & Reents, R. 2007. Validation of an approximate REML algorithm for parameter estimation in a multitrait, multiple across-country evaluation model: A simulation study. *J. Dairy Sci.* 90, 4846-4855.

Table 2. Estimates of genetic correlations between the selected traits. Production and SCS traits are on combined lactation basis.

	FY	PY	SCS	L	FSh	NRh	CF	NRc	FSc	DO	BD	RA	UD	OD	FL	LO	BCS ¹
Milk (MY)	.39	.88	.04	.09	.20	.02	.35	-.04	.37	.45	.04	.04	-.24	.06	.05	.00	-.27
Fat (FY)		.63	-.01	.08	.16	-.03	.24	-.08	.30	.33	.09	-.01	-.18	.01	.03	.02	-.14
Protein (PY)			.04	.07	.19	-.02	.30	-.08	.36	.41	.05	.04	-.24	.05	.08	.04	-.23
Somatic cell scores (SCS)				.59	.06	-.08	.20	-.01	.11	.19	.15	.05	-.32	-.20	-	-	-.11
Longevity (L, as relative risk)					.21	-.07	.54	-.01	.40	.58	.34	-.05	-.48	-.37	-	-	-.13
Interval first to successful insemination virgin heifer (FSh)											.11	-.03	-.09	-.00	-	-	-.09
Non-return rate 56 days virgin heifer (NRh)															.01	.03	
Interval calving to first insemination (CF)											-.13	-.02	.03	.02	.02	.09	-.07
Non-return rate 56 days cow (NRc)											.09	.03	-.22	-.06	-	-	-.25
Interval first to successful insemination cow (FSc)															.15	.17	
Days open (DO)											-.12	.07	.00	-.03	-	-	-.05
Body depth (BD)											.19	-.05	-.16	-.00	-	-	-.18
Rump angle (RA)															.08	.05	
Udder depth (UD)															.02	.06	
Overall udder score (OD)											.17	-.01	-.23	-.04	-	-	-.27
Overall feet and legs score (FL)															.10	.14	
Locomotion (LO)																.03	.27
Body condition score (BCS)																-	-.03
																.08	
																.09	.02
																.29	-.05
																.74	.08
																	.15

¹ BCS stands for body condition score.

Table 3. Estimates of genetic correlations of milk yield and SCS in first three lactations.

	Milk yield in lactation			SCS in lactation		
	First	Second	Third	First	Second	Third
Longevity, as relative risk	.26	.08	-.06	.48	.59	.61
Interval first to successful insemination virgin heifer (FSh)	.19	.18	.18	.07	.05	.06
Non-return 56 virgin heifer (NRh)	-.01	.02	.04	-.08	-.07	-.09
Interval calving to first insemination (CF)	.36	.34	.29	.20	.19	.19
Non-return 56 cow (NRc)	-.03	-.03	-.04	-.02	.00	-.01
Interval first to successful insemination cow (FSc)	.35	.35	.34	.13	.10	.10
Days open (DO)	.44	.42	.39	.20	.18	.18
Body depth (BD)	.08	.03	.01	.12	.14	.16
Rump angle (RA)	.07	.03	.02	.03	.04	.05
Udder depth (UD)	-.32	-.21	-.16	-.28	-.31	-.32
Overall udder score (OD)	.01	.06	.07	-.20	-.19	-.19
Overall feet and legs score (FL)	-.00	.05	.08	-.07	-.06	-.06
Locomotion (LO)	-.05	.02	.03	-.11	-.09	-.07
Body condition score (BCS)	-.31	-.24	-.21	-.14	-.10	-.09

Figure 1. Estimated genetic correlations between milk yield and somatic cell scores

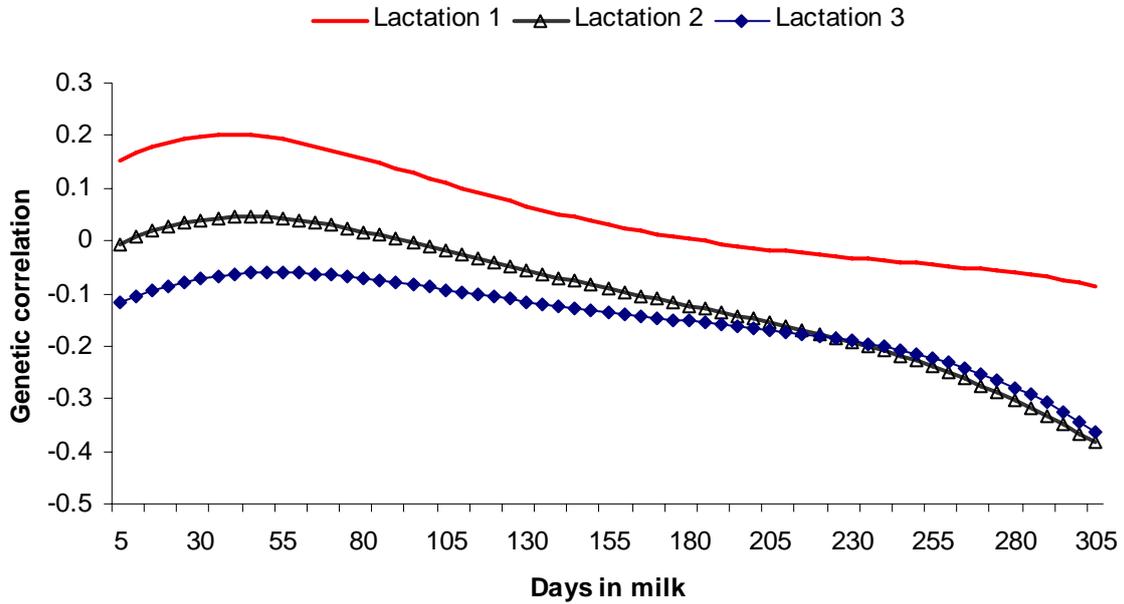


Figure 2. Estimated genetic correlations between milk yield and fat yield

