A stochastic simulation study for the comparison of different methods to calculate a total merit index

C. Pfeiffer[1], B. Fuerst-Waltl[1], H. Schwarzenbacher[2], F. Steininger[2], C. Fuerst[2]

[1]University of Natural Resources and Life Sciences, Department of Sustainable Agriculture, Division of Livestock Sciences
Vienna, Gregor-Mendel-Straße 33, 1180 Vienna, Austria
[2]ZuchtData EDV-Dienstleistungen GmbH, Dresdner Straße 89/19, 1200 Vienna, Austria

ABSTRACT

In order to compare four different methods for calculation of a total merit index, a stochastic simulation study was conducted. Five normally distributed traits were chosen to represent the blocks dairy, beef and fitness of a simulated cattle population. The reference method was a full multivariate evaluation based on raw data. The other three methods were based on selection index theory with different approaches to calculate covariances between estimated breeding values. Additionally a focus was put on the implications of varying the residual covariances between traits. All selection index methods showed similar results. However, the method currently used in the joint genetic evaluation led to noticeable biases in EBVs especially when residual covariances between traits were high. Residual covariances seem to have an important impact when calculating a total merit index and should not be ignored. Results of the present study encourage to move towards a multitrait approach or at least to account for residual covariances when combining EBVs into a total merit index.
(Keywords: total merit index, multitrait evaluation, stochastic simulation)

INTRODUCTION

The total merit index (TMI), which is a function of different estimated breeding values (EBV), is used as one of the most important selction criterions worldwide (*Miglior et al., 2005*). In modern dairy cattle breeding programs, the TMI is commonly based on different production and increasingly on functional traits. Typically EBVs of different traits are weighted, concerning their economic importance and combined to a TMI (*Hazel and Lush*, 1942). The use of selection index theory is however faced with some challenges: Traits or group of traits are usually evaluated separately based on different statistical models, and hence true genetic or/and phenotypic correlations or heterogeneous reliabilities are neglected (*Ducrocq et al.*, 2001). This is also the case in the joint genetic evaluation of Austrian and German dairy cattle breeds. The TMI and several sub-indices for all cattle breeds except Holstein is based on a selection index method (*Hazel and Lush*, 1942) which was proposed by *Miesenberger* (1997). The TMI of Fleckvieh (dual purpose Simmental) and Brown Swiss currently consists of more than 20 different production and functional traits. EBVs for the TMI as well as for several sub-indices are estimated either univariately or multivariately in different linear or non linear models. Subsequently EBVs are combined to TMIs or to other sub-indices assuming that residual covariances between traits or group of traits are zero. A full

multivariate estimation of all traits based on raw data could be considered as the optimum methodology but is usually not feasible (*Mrode*, 2014). Although computer power and capacity is increasing quickly it is still demanding to compute all traits, which are included in a TMI, together. Experiences of the last years suggest that particularly TMIs with low reliabilities (r^2) are slightly overestimated. Much effort is put on an approximate two-step procedure (*Fuerst et al.,* 2014; *Pfeiffer et al.,* 2014), which was proposed by *Ducrocq et al.* (2001) and validated by *Lassen* et al. (2007). However, alternative combinations of independently estimated breeding values are also evaluated. Apart from the method proposed by *Miesenberger* (1997), two additional similar methods described by *Götz* (2002) are still in discussion. Hence, the objective of the present study was the comparison of these methods with a full multitrait animal model. This was done in a stochastic simulation study mimicking a simplified breeding scheme of Austrian Brown Swiss cattle. Special attention was also put on assuming different residual covariances.

MATERIAL AND METHODS

A population structure roughly reflecting the Austrian Brown Swiss cattle population was simulated with the stochastic simulation program ADAM (*Pedersen et al.,* 2009). Approximately 51,300 cows distributed on 1,710 herds were simulated. Five traits following Gaussian distribution were chosen to represent the blocks dairy (fat (FY) and protein (PY) yield), beef (net daily gain (NDG)) and fitness (somatic cell count (SCC) and non-return rate (NRR) of cows). Further requirements were a wide range of heritabilities and genetic correlations as well as economic importance. Four traits FY, PY, SCC and NRR were measured on all female animals, NDG was observed on approximately 60% of all male animals. Each trait was measured on every animal in all herds, no repeated records were assumed. The assumed heritabilities and genetic correlations for the five traits are shown in *Table 1*. Around 25% of young bulls and 75% of proven bulls were used for matings in the selection scheme. Breeding values and phenotypes for the five traits were simulated for base population animals. Afterwards animals were selected on a TMI based on multivariately estimated breeding values (EBV) over 30 years. Relative economic weights for FY, PY, NDG, SCC and NRR were adopted from the values used in routine genetic evaluation, which are 5.4, 53.6, 4.3, 19.7 and 17% respectively (*Fuerst et al.,* 2013). Three different scenarios with respect to the covariances of the residual effects were simulated. In scenarios 0, 1 and 2 residual correlations were varied from zero, to half and equal to the genetic correlations, respectively. The variation of the residual covariances was specifically evaluated to appraise the impact of ignoring residual covariances. In total, ten replicates were conducted for each scenario.

Table 1

Heritabilities (on the diagonal) and true genetic correlations (above diagonal)

Trait	FY	PY	NDG	SCC	NRR
FY	0.40	0.85	0.10	0.25	-0.20
PY		0.39	0.10	0.25	-0.20
NDG			0.27	0.00	0.00
SCC				0.12	-0.10
NRR					0.02

Method A was a full multitrait animal model based on raw data using the true genetic and phenotypic parameters. The model included a fixed herd-year-effect, a random genetic and a random residual effect. Subsequently the TMI was calculated as:

$$TMI_A = EBV_{FY}\,\omega_{FY} + EBV_{PY}\,\omega_{PY} + EBV_{NDG}\,\omega_{NDG} + EBV_{SCC}\,\omega_{SCC} + EBV_{NRR}\,\omega_{NRR} \quad (1)$$

where **EBV** refers to the certain traits; ω denotes the relative economic weights which are 5.4% for FY, 53.6% for PY, 4.3% for NDG, 19.7% for SCC and 17% for NRR, respectively. Method A was considered to be the reference method. For methods B, C and D, EBVs were estimated in univariate animal models including the same effects described above. In order to obtain the TMI of method B, which is the currently used method (proposed by *Miesenberger,* 1997), C (proposed by *Dempfle; Götz*, 2002) and D (proposed by *Reinhardt*; *Götz*, 2002) equation (1) was applied and covariances between the EBVs (σ_{ij}) of the different methods (indicated by sub-indices B, C, D) were calculated as:

$$\sigma_{ijB} = r_{gij}\,r^2_i\,r^2_j\,\sigma_{ai}\,\sigma_{aj} \quad (2)$$

$$\sigma_{ijC} = r_{pij}\,r_i\,r_j\,\sigma_{ai}\,\sigma_{aj} \quad (3)$$

$$\sigma_{ijD} = r_{gij}\,r_i\,r_j\,\sigma_{ai}\,\sigma_{aj} \quad (4)$$

where r_{gij} is the genetic correlation between traits i and j; $r^2_{i,j}$ are the reliabilities of EBVs of traits i and j; $\sigma_{ai,j}$ are the additive genetic standard deviations of traits i and j; $r_{i,j}$ are the accuracies of EBVs of trait i and j and r_{pij} is the phenotypic correlation between traits i and j.
This means that only method C accounts for residual correlations.

Estimated breeding values were calculated using the program package MiX99 (*Lidauer et al.,* 2013). For all methods, genetic parameters were not re-estimated. The true (simulated) simulated parameters were used. All EBVs were standardised to 12 points per additive genetic standard deviation. The base was set to 100 for the years 18 to 22.

RESULTS AND DISCUSSION

Across all year groups Spearman rank correlations between the true and the estimated breeding values were about 0.86 for scenario 0 and about 0.83 for scenario 2. For scenario 1, which is not shown in *Table 2*, the correlation across all year groups is above 0.86.

As in scenario 2 genetic and phenotypic correlations are identical, the results for methods C and D are the same. Rank correlations between true and estimated TMIs across year groups are moderate, because of relatively low reliabilities in the simulated population (approximately 41% of the simulated animals have a r^2 below 60%). Rank correlations within year groups are rather similar, but slightly lower for method B in scenarios 1 and 2. Rank correlations of scenario 2 are in general slightly lower than the correlations of scenarios 0 and 1. Furthermore rank correlations between the full multivariate method (A) and all other methods, including all scenarios were calculated. Rank correlations are in the range of 0.93 to 0.99 within year groups. Across all animals rank correlations are between 0.98 and 0.99.

Table 2

Rank correlations between the true TMI within year groups for different methods for scenarios 0 and 2

Scenario	Years	A	B	C	D
0	All	0.8704	0.8612	0.8606	0.8620
	11–15	0.6399	0.6149	0.6054	0.6122
	16–20	0.6516	0.6283	0.6244	0.6258
	21–25	0.6262	0.5923	0.5953	0.5980
	26–30	0.6657	0.6343	0.6406	0.6404
Scenario	Years	A	B	C	D
2	All	0.8490	0.8346	0.8476	0.8476
	11–15	0.6274	0.5948	0.6233	0.6233
	16–20	0.6237	0.5890	0.6196	0.6196
	21–25	0.5790	0.5288	0.5745	0.5745
	26–30	0.6190	0.5785	0.6151	0.6151

In this study biases are products of subtracting the true TMI from the estimated TMI. This was done for all animals and scenarios. *Table 3* shows the bias of scenarios 0 and 2. Results for scenario 1 are between scenario 0 and 2.

Table 3

Bias of different TMI methods from the true TMI within year groups for scenarios 0 and 2

Scenario	Years	A	B	C	D
0	All	-0.1	0.1	0.1	0.2
	11-15	-0.4	-1.5	-1.2	0.4
	16-20	-0.1	-0.4	-0.4	0.1
	21-25	0.1	0.7	0.6	0.1
	26-30	0.1	1.5	1.2	0.1
2	All	-0.1	-0.6	-0.1	-0.1
	11-15	-0.3	-3.0	-0.7	-0.7
	16-20	-0.1	-0.8	-0.2	-0.2
	21-25	0.0	0.8	0.2	0.2
	26-30	-0.1	0.7	0.2	0.2

Results of scenario 0, where no residual covariances were assumed, show very good results particularly for methods A and D. Methods B and C seem to underestimate the animals in the first years in both scenarios. One possibility can be an incomplete pedigree and the use of phantom parents groups (*Fuerst et al.,* 2014). However, method B leads to an overestimated genetic trend. This trend is more pronounced when residual covariances are assumed. This overestimation is even stronger in the best 10% animals in TMI per year. *Table 4* shows the bias (EBV-TBV) of the TMIs of the top 10% animals within year groups.

Figure 1 shows the bias for the top animals in scenario 1, which is expressed as a downwards bias in the first years and an upwards bias in the last years.

Table 4

Bias of different TMI methods from the true TMI for the top 10% within year groups for scenarios 0 and 2

Scenario	Years	A	B	C	D
0	All	0.2	1.5	1.7	1.0
	11-15	-0.2	-0.6	0.3	1.0
	16-20	0.1	0.4	1.0	0.9
	21-25	0.4	2.2	2.3	0.9
	26-30	0.5	3.8	3.1	1.0
2	All	0.2	2.3	0.6	0.6
	11-15	-0.2	-0.4	-0.2	-0.2
	16-20	0.2	1.7	0.5	0.5
	21-25	0.4	3.7	1.0	1.0
	26-30	0.4	4.1	1.0	1.0

Figure 1

Time trend of bias (EBV-TBV) of different methods for the top 10% animals within years for scenario 1
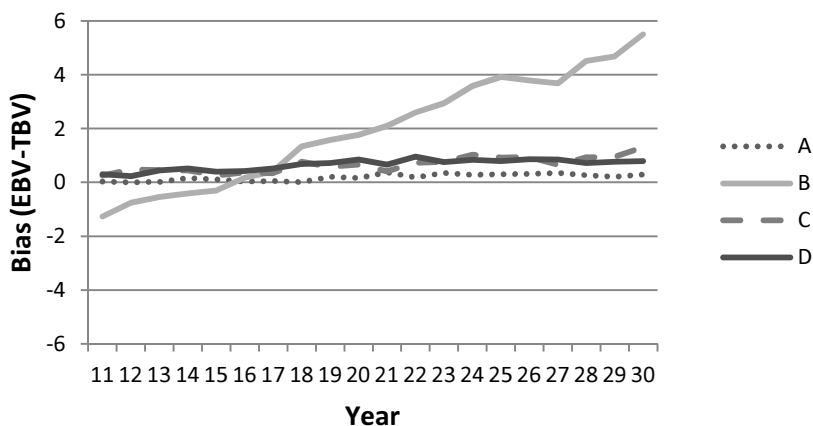
CONCLUSIONS

Results show that all methods based on selection index theory are quite similar. The analysed methods show good results when residual covariances are zero. However, in real data residual covariances can have an important impact. It is well known that omitting residual covariances when the same animals are recorded in the same environment is not valid. The currently used method B shows good results for high

reliabilities but leads to inflated deviations mainly in case of low reliabilities. This results in a bias particularly for the top animals and can therefore be relevant in terms of selection accuracy. For the joint genetic evaluation of Austria and Germany, it is intended to replace the current method of TMI calculation by a multitrait approach. If this is not working, an adapted method of including residual covariances between traits is needed.

ACKNOWLEDGEMENTS

REFERENCES

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. In Proceedings of the 52nd annual meeting EAAP, Budapest, Hungary; 2001.

Fuerst, C., Dodenhoff, J., Egger-Danner, C., Emmerling, R., Hamann, H., Krogmeier, D., Schwarzenbacher, H. (2013). Zuchtwertschätzung beim Rind. http://www.zar.at/download/ZWS/ZWS.pdf.

Fuerst, C., Pfeiffer, C., Schwarzenbacher, H., Steininger, F., Fuerst-Waltl, B. (2014). Comparison of different methods to calculate a total merit index – results of a stochastic simulation study. Manuscript accepted for the 10th World Congress on Genetics Applied to Livestock Production, August 17-22, 2014, Vancouver, BC, Canada.

Götz, K.-U., (2002). Optimale Kombination von unabhängig voneinander geschätzten Zuchtwerten. Unpublished document.

Hazel, L.N, Lush, J.L (1942). The efficiency of three methods of selection. Journal of Heredity, 33. 393-399.

Lidauer, M., Matilainen, K., Mäntysaari, E., Pitkänen, T., Taskinen, M., Strandén, I. (2013). MiX99.General program for solving large mixed model equations with preconditioned conjugate gradient method. Relase VII/2013 beta, Jokioinen, Finland.

Lassen, J., Sørensen, M.K., Madsen, P., Ducrocq, V. (2007). A stocastic simulation study on validation of an approximate multitrait model using preadjusted data for prediction of breeding values. Journal of Dairy Science, 90. 3002-3011.

Miesenberger, J. (1997). Zuchtzieldefinition und Indexselektion für die österreichische Rinderzucht. PhD thesis, University of Natural Resources and Life Sciences Vienna, Austria.

Miglior, F., Muir, B.L., Van Doormaal, B.J. (2005). Selection indicies in Holstein cattle in various countries. Journal of Dairy Science, 88. 1255-1263.

Mrode, R.A. (2014). Linear models for the prediction of animal breeding values, 3nd edition, CAB International.

Pedersen, L.D., Sørensen, A.C., Henryon, M., Ansari-Mahyari, S., Berg, P. (2009). ADAM: A computer program to simulate selective breeding schemes for animals. Livestock Science, 121. 343-344.

Pfeiffer, C., Fuerst-Waltl, B., Ducrocq, V., Fuerst, C. (2014). Approximate multivariate genetic evaluation of functional longevity and type traits in Austrian Fleckvieh cattle. Manuscript accepted for the 10th World Congress on Genetics Applied to Livestock Production, August 17-22, 2014, Vancouver, BC, Canada.

Corresponding author:

Christina Pfeiffer
University of Natural Resources and Life Sciences
Department of Livestock Sciences
Gregor-Mendel Strasse 33, A-1180 Vienna
Phone: +43-1-47-654-3252
E-mail: christina.pfeiffer@boku.ac.at