Improved genetic evaluation of survival using MACE to combine direct and correlated information from yield and functional traits

Tom Druet[1], Johann Sölkner[2], Ab F. Groen[3] and Nicolas Gengler[1]
[1]*Animal Science Unit, Gembloux Agricultural University, B-5030 Gembloux, Belgium,*
[2]*Department of Livestock Science, University of Agriculture, Gregor-Mendel-Strasse 33; A-1180 Wien, Austria*
and
[3]*Animal Breeding and Genetics Group, Wageningen Institute of Animal Sciences, Wageningen Agricultural University, P.O. Box 338, 6700 AH Wageningen, The Netherlands*

Abstract

Objective of this study was to improve the estimation of survival breeding values using multiple-trait across country evaluation procedures to combine direct information with milk, fat and protein yields and functional traits: milk persistency, somatic cell count, fertility (male and female) and calving ease (direct and maternal). A set of bulls was selected from the Austrian Simmental population and their breeding values were deregressed. An Expectation-Maximization REML procedure based on the multiple-trait across country evaluation equations was used to compute genetic (co)variance components among all traits and to predict breeding values. Mean rank correlation between direct and combined breeding values was 0.85. Direct (original), indirect and combined reliabilities were compared to appreciate the impact of the proposed method on improvement of reliability of survival breeding values. This improvement could be up to 0.24 for animals with low direct reliability. For young bulls born in 1992 without direct survival information but with already reliable information on milk, fat and protein yields and functional traits, mean reliability improvement was 0.13. For all bulls this value was 0.06.

1. Introduction

Since several years many studies (e.g. Congleton & King, 1967) have shown that selection for longevity has a positive economic impact. But there are two major obstacles for selecting on such a trait: low heritability and late availability (e.g. Smith, 1983). Ducrocq (1987) proposed survival analysis as a method to get earlier information by the optimal use of censored (not yet finished) records. However the breeding value of a bull when his daughters are in the end of their first lactation, is still not very reliable as the number of daughter that are actually culled is limited.

The multiple-trait across country evaluation (MACE) procedure was developed for international sire comparison (Schaeffer, 1993; 1994). Weigel (1996) pointed to the similarities in the use of information from correlated traits to improve the accuracy of prediction of breeding values for longevity (herd life) as suggested earlier (e.g. Boldman *et al.*, 1992) and the international evaluation of dairy sires: sire evaluation data from correlated traits from different sources are combined in an optimal manner. Weigel (1996)

suggested therefore the use of the MACE to improve early prediction of productive life. MACE can be adapted for multiple a posteriori genetic evaluation for different traits in the same country under the assumption that breeding values were obtained on different sets of daughters. The accuracy of the prediction will be higher because a multiple-trait model combines more information from several sources (i.e. traits). In this way, the use of MACE procedure could reduce some problems of the selection for herd life, productive life or survival as it could improve the accuracy of early predictions. Jairath *et al.* (1998) applied MACE procedures to improve genetic evaluation for herd life in Canada by using information on milk production traits and conformation traits. The objective of this study was to use MACE approaches to combine survival values with milk production traits and functional traits in the Austrian (dual purpose) Simmental population. A second objective was to estimate needed (co)variance components also using MACE.

2. Material and methods

2.1. Data

The data used for this study was provided by the Zentrale Arbeitsgemeinschaft österreichischer Rinderzüchter. The data consisted of the breeding values available in January 1998 of 564868 Austrian Simmental bulls. Ten traits were considered: milk, fat and protein yields, survival, milk persistency, somatic cell count, fertility (female and male) and calving ease (maternal and direct). Bulls selected were born between 1985 and 1995 and had a mean reliability of their official breeding values for the yield traits higher or equal to 0.50 and a mean reliability of their official breeding values for the functional traits higher or equal to 0.20. The male ancestors in the pedigrees of these 6918 bulls were also included in the estimations. A total of 9712 bulls were involved in this study.

2.2. Methods

Four steps were involved in this study:
− estimation of daughter equivalents;
− deregression of sire breeding values;
− estimation of sire variances and genetic correlations with an Expectation-Maximization REML procedure;
− evaluation of new breeding values and their reliabilities by a multiple-trait analysis.

Number of daughters were not available therefore the concept of daughter equivalents was used (VanRaden & Wiggans, 1991) including also information from relatives as suggested by Schaeffer (1993):

$$DE = REL(1 - REL)^{-1} (4 - h^2)/h^2 \quad [1]$$

where DE = daughter equivalent, REL = original reliability (defined between 0 and 1) and h^2 = heritability used in the computations. The correct weighting of DYD or deregressed proofs should be by the inverse of their relative residual variances and the values obtained by [1] are obviously only approximates.

MACE programs were provided by B. Klei (Holstein Association of America, Brattleboro) and G. Banos (INTERBULL Center, Uppsala). These programs deregressed proofs, computed genetic correlations between all traits and computed solutions of the multiple trait MACE model. MACE models require daughter yield deviation (DYD) of the bulls. If these are not available, it is possible to deregress the proofs to get similar information (Banos et al., 1993).

The deregression was realized by an iterative process accounting for the mean and also including phantom parent groups as described by Jairath et al. (1998), pretending that the breeding values were obtained by single trait sire-maternal grand sire models. The aim of the deregression is to regenerate the right-hand side of the mixed model equations. These regenerated values with the fixed effects absorbed and adjusted for the mean are a measurement of the DYD (Lien et al., 1995) or the corrected mean deviation of a daughter group. Each trait was deregressed separately.

Deregressed values from all the traits were combined and the Expectation-Maximization REML algorithm described by Sigurdsson & Banos (1995) was used to estimate the genetic correlations between all these traits.

The model used was exactly the classic MACE model by Sigurdsson & Banos (1995):

$$\mathbf{y} = \mathbf{Xc} + \mathbf{ZQg} + \mathbf{Zs} + \mathbf{e} \quad [2]$$

where $\mathbf{y}$ = vector of deregressed proofs; $\mathbf{c}$ = vector of fixed effects for each trait (mean); $\mathbf{g}$ = vector of genetic group effects; $\mathbf{s}$ = vector of random bull effects; $\mathbf{X}$ = incidence matrix linking $\mathbf{y}$ and $\mathbf{c}$, $\mathbf{Z}$ = incidence matrix linking $\mathbf{y}$ and $\mathbf{s}$, $\mathbf{Q}$ = matrix assigning bulls to phantom parent groups and $\mathbf{e}$ = vector of random residual effects.

Mixed model equations associated with model [2] were:

$$\begin{bmatrix} \mathbf{X'R^{-1}X} & \mathbf{0} & \mathbf{X'R^{-1}Z} \\ \mathbf{0} & \mathbf{Q'A^{-1}Q \otimes G^{-1}} & \mathbf{-Q'A^{-1} \otimes G^{-1}} \\ \mathbf{Z'R^{-1}X} & \mathbf{-A^{-1}Q \otimes G^{-1}} & \mathbf{Z'R^{-1}Z + A^{-1} \otimes G^{-1}} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{c}} \\ \hat{\mathbf{g}} \\ \mathbf{Q\hat{g}+\hat{s}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'R^{-1}y} \\ \mathbf{0} \\ \mathbf{Z'R^{-1}y} \end{bmatrix}$$

where $\mathbf{R}$ = diagonal matrix of residual variances, $\mathbf{A}$ = additive relationship matrix among bulls and $\mathbf{G}$ = matrix of genetic variance-covariances among traits, $\otimes$ is the Kronecker operator. It must be again stressed here that pretending that $\mathbf{R}$ being diagonal is only an approximation, but Jairath et al. (1998) deducted from selection index results that impact of this assumption should be relatively low.

Approximately 3500 animals and 6 traits could be used at the same time to estimate variance components. Therefore four samples were drawn from the 6918 bulls born between 1985 and 1995. The bulls were sorted by identification number and we choose one bull of four. Afterwards the male ancestors in the pedigrees were added to every sample. The size of the samples were between 3358 and 3415 bulls.

Four groups of traits were created: milk, fat and protein yields; survival, milk persistency and somatic cell count; paternal or direct components of fertility and calving ease; maternal components

of the same traits. All these groups were combined two by two resulting in 6 runs per sample therefore in total 24. For a sample the genetic correlations matrix of the ten traits was built by calculating the mean correlations from the 6 runs among all the traits. These correlation matrices were positive definite. Results for the four samples were then averaged. Empirical standard errors were computed as observed standard deviation among the four samples. At the end, the mean variances obtained by a similar procedure were used to compute the variance-covariance matrix necessary for the computation of the MACE solutions.

The MACE equations were solved by an iterative Gauss-Seidel approach as described by Klei (1998).

In order to approximate multitrait reliabilities the following method was used. Formula [1] can be rewritten for all the traits as:

$$\mathbf{D} = \mathbf{C}_{st} \, (\mathbf{I} - \mathbf{C}_{st})^{-1} \mathrm{diag}\{\mathbf{G}\}^{-1}\mathbf{R} \quad [3]$$

where $\mathbf{D}$ = diagonal matrix of daughters equivalents, $\mathbf{C}_{st}$= diagonal matrix of single-trait reliabilities, $\mathrm{diag}\{\mathbf{G}\}$ = matrix of diagonal elements of the genetic (co)variance matrix of sire effects and $\mathbf{R}$ = diagonal matrix of residual variances. Multitrait reliabilities can then be approximated by estimating a matrix $\mathbf{P}$ of multitrait prediction error variances (PEV) (Harris & Johnson, 1998). For a given animal PEV is:

$$\mathbf{P} = \mathrm{diag} \, \{(\mathbf{DR}^{-1} + \mathbf{G}^{-1})^{-1}\} =$$
$$\mathrm{diag} \, \{(\mathbf{C}_{st}(\mathbf{I} - \mathbf{C}_{st})^{-1}\mathrm{diag}\{\mathbf{G}\}^{-1}+ \mathbf{G}^{-1})^{-1}\} \quad [4]$$

Multitrait reliabilities are then approximated as:

$$\mathbf{C}_{mt} = \mathbf{I} - \mathbf{P} \, (\mathrm{diag} \, \{\mathbf{G}\}^{-1}) \quad [5]$$

$$\mathbf{C}_{mt}=\mathbf{I}-\mathrm{diag}\{(\mathbf{C}_{st}(\mathbf{I}-\mathbf{C}_{st})^{-1}\mathrm{diag}\{\mathbf{G}\}^{-1}+\mathbf{G}^{-1})^{-1}\}\mathrm{diag}\{\mathbf{G}\}^{-1}$$
$$[6]$$

where $\mathbf{C}_{mt}$ = diagonal matrix of multitrait reliabilities.

This method does consider animals unrelated what is not necessarily a bad approximation as parent and progeny information are already contained in $\mathbf{C}_{st}$.

Two types of multitrait reliabilities were obtained: indirect, considering only the contributions from correlated traits, and combined, considering both direct and indirect contributions.

3. Results & discussions

3.1.Genetic correlations

Most of studies, estimated genetic correlations between survival and conformation traits. Here, we had the opportunity to estimate correlations with functional traits. The correlations among the ten traits are presented in Table 1 with empirical s.e. (standard errors) estimated as the s.e. of the four sub-samples. Empirical s.e. were in general low with values close to 0.01 or even lower except for some correlations involving functional traits with low heritabilities (such as fertility or calving ease traits) where the empirical s.e. were up to 0.06.

Correlations of survival with yield traits were small but still between 0.10 and 0.15. Theoretically these values should be close to zero as functional survival is (phenotypically) corrected for yields on a within herd basis. Correlations of survival with milk persistency, maternal fertility and calving ease traits were all above 0.20 and this in a very consistent pattern across samples. The most correlated trait seemed to be maternal fertility with a value over 0.30. Miesenberger et al. (1998) stressed already the fact that in the Simmental breed, more cows are culled for fertility disorders than for low milk production. Interesting is also the correlation of 0.23 with milk persistency. Correlations with calving ease traits were as expected as difficult calvings are lowering the chances of survival. A recent study (Druet, 1998) using less reliable samples with fewer bulls for Austrian breeds, notably Holstein and Brown Swiss, showed that correlations between survival and maternal fertility, milk persistency and somatic cell count ranged from 0.09 to 0.49 and that correlations between survival and milk production traits were variable, as already noticed by Ducrocq (1987) or Vollema & Groen (1997), nearly zero for Brown Swiss and negative for Holstein. Somatic cell count had a positive relationship with functional survival, with the strongest correlation in Holstein breed. This confirms results by Rogers et al. (1996) who estimated correlations of 0.06 and 0.30 between somatic cell count and herd life, and by Jairath et al. (1998), who estimated a correlation of 0.17 among the same traits.

Functional traits and conformation traits are complementary because they are related to different culling reasons such as fertility disorders, mastitis, dystocia or feet and legs problems. In addition, milk persistency seems to be a usefull predictor. Some other characters could be added such as milkability and behaviour.

In Holstein, conformation traits related to the mammary system seem to show a high correlation with survival related traits, around 0.30 (Vollema & Groen, 1997; Jairath et al., 1998). In this same breed, comparable high correlations were also found between somatic cell count and survival so that both conformation and functionals traits

related to mastitis could be used as survival predictors. Feet and legs conformation traits are also usefull in order to predict survival abilities because their are directly related to an important culling reason in some breeds. Therefore their correlation with survival are also around 0.30 (Jairath et al., 1998). Another conformation trait who showed often a high correlation with survival is general or final score (Vollema & Groen, 1997; Jairath et al., 1998) but we have to remember that the effect of such a trait on survival is heavily dependent of the breed and on the fact that the cow is registered or not (Vollema & Groen, 1997; Ducrocq, 1987). Rump width conformation traits were lesser correlated (e.g. around 0; Weigel et al., 1998; Jairath et al., 1998) with survival than maternal calving ease but comparison should be done with the same breed.

Table 1. Genetic correlations between milk production traits and functional traits in Austrian Simmental cattle and their empirical standard errors.

	Fat	Protein	Survival	Persistency	Somatic cell count[1]	NRR90 (p)[2]	NRR90 (m)[3]	Calving ease (d)[4]	Calving ease (m)[5]
Milk	**0.78** ±0.00	**0.91** ±0.00	**0.11** ±0.02	**0.06** ±0.02	**-0.07** ±0.03	**-0.07** ±0.04	**-0.22** ±0.03	**0.03** ±0.07	**-0.02** ±0.02
Fat		**0.83** ±0.00	**0.14** ±0.03	**0.08** ±0.02	**-0.01** ±0.01	**-0.03** ±0.06	**-0.16** ±0.04	**0.08** ±0.06	**0.01** ±0.05
Protein			**0.14** ±0.03	**0.11** ±0.02	**-0.04** ±0.02	**-0.06** ±0.05	**-0.18** ±0.04	**0.05** ±0.06	**0.00** ±0.01
Survival				**0.23** ±0.03	**0.11** ±0.02	**-0.08** ±0.05	**0.34** ±0.02	**0.21** ±0.01	**0.22** ±0.03
Persistency					**0.17** ±0.01	**0.05** ±0.04	**0.13** ±0.05	**0.14** ±0.03	**0.12** ±0.03
Somatic cell count[1]						**-0.02** ±0.05	**0.01** ±0.01	**0.05** ±0.02	**-0.01** ±0.03
NRR90 (p)[2]							**-0.16** ±0.02	**0.11** ±0.05	**-0.01** ±0.01
NRR90 (m)[3]								**0.00** ± 0.04	**0.19** ±0.03
Calving ease (d)[4]									**-0.05** ±0,03

[1] For this character higher values are desirable; [2] NRR90 (p) = paternal component of non-return rate at 90 days; [3]NRR90 (m) = maternal component of non-return rate at 90 days; [4] Calving ease (d) = direct component of calving ease score; [5] Calving ease (m)= maternal component of calving ease score.

3.2. Prediction of survival

Rank correlations between direct and combined breeding values for survival are in Table 2. Mean correlation between official breeding values and the newly estimated ones are 0.85, close to the result obtained by Jairath et al. (1998) who estimated a correlation of 0.80 between direct and combined herd life but using conformation information.

The rank correlations showed a higher influence of MACE procedures on the ranking for young bulls. Rank correlation was 0.88 for bulls born in 1987 and only 0.67 for the bulls born in 1992. For these last bulls, Figure 1. shows the direct proofs and the estimated breeding values (MACE solutions). There was a clear re-ranking between these 34 bulls and using MACE solutions would have had some influence of selection decisions.
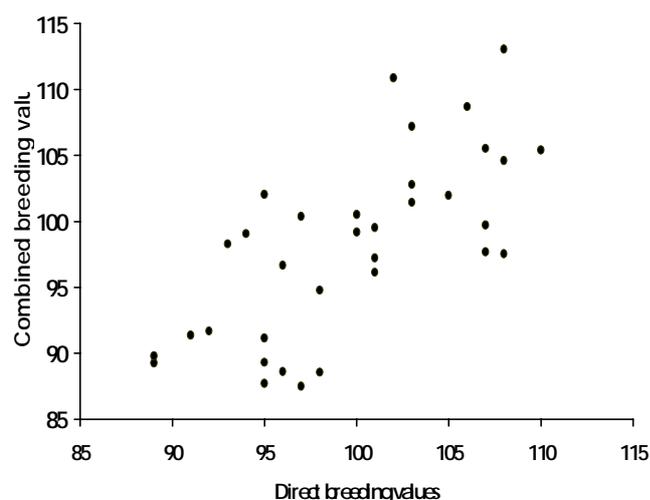

Figure 1. Comparison of estimated breeding values for survival from multiple trait across countries evaluation (MACE) and direct breeding values for bulls born in 1992.

It was checked if these re-ranking could be due only to change in reliability of the proofs and a negative answer was found.

3. 3. Direct, indirect and combined reliabilities

Direct or original, indirect obtained from correlated traits and combined reliabilities are in Table 2. In order to compare young bulls to older, results are given by birth year. Results confirmed the expectations that reliability improvements were rather important for younger animals. For bulls born in 1992, estimated breeding values were very inaccurate, basically parent averages, but indirect reliability was nearly as high as direct therefore

combined reliability was 0.13 reliability points (or 50 %) higher than direct reliability.

The most affected animals were those for which direct reliability values were low and indirect information was high. The improvement could be up to 0.24 reliability points or nearly 100 % for such animals. For all the animals, mean gain of reliability was 0.06 which represents still 12 % of their original mean reliability.

Weigel *et al*. (1998) computed reliabilities for direct, indirect and combined predictions of productive life with American Holstein. Predictors were milk production traits and conformation traits. Mean reliabilities for direct, indirect and combined predictions were respectively 0.50, 0.33 and 0.52. These values are similar to those we obtained with functional traits.

Table 2: Rank correlations between direct and combined breeding values by birth year, and direct, indirect and combined reliabilities by birth year.

Year of birth	Number of Bulls	Rank correlation	Reliability		
			Direct	Indirect	Combined
before 1985	1552	0.868	0.64	0.33	0.68
1985	458	0.839	0.42	0.23	0.49
1986	370	0.838	0.40	0.21	0.47
1987	385	0.877	0.39	0.21	0.46
1988	313	0.863	0.38	0.22	0.46
1989	276	0.813	0.38	0.22	0.46
1990	210	0.701	0.38	0.24	0.47
1991	117	0.696	0.38	0.26	0.49
1992	34	0.674	0.26	0.23	0.39
All the bulls	3715	0.854	0.50	0.27	0.56

4. Conclusions

MACE procedures offer the possibility to improve early prediction of survival breeding values by using information on correlated traits. Some traits are indicated to improve accuracy of survival values. The most important ones are in a maternal fertility trait, milk persistency, a trait related to mastitis problems as somatic cell count or a udder conformation trait, a feet and legs conformation trait and calving ease traits. Milk production traits have also to be included in such a model. Milking and behaviour traits could also bring additional information.

However, current MACE procedures are only approximate pretending that non-genetic correlations among DYD or deregressed proofs are zero. New ideas are needed to develop procedures which avoid this simplification. Also deregression procedures showed some limits for traits with poor heritability and animals with poor reliabilities. A possible improvement would be to avoid deregression by using DYD rather than deregressed proofs.

We have also to consider the impact of genetic correlations on indirect and combined reliability. The higher these correlations are, the greater will be the influence of the correlated traits. Unfortunately estimation of correlations is always rather risky and we need to acknowledge that

different correlations might change results. If we choose too high correlations, we overestimate the influence of other traits and this can lead to some bias but if we limit these genetic correlations, we lose information through correlated traits. Also there are indications that correlations have to be estimated for each breed in each country and at a particular moment. Despite these points our results showed that the use of MACE can be a considerable step forward to improve early prediction of survival breeding values.

References

Banos, G., Philipsson, J., Gundel, M., Juga, J. and Sander-Nielsen, U. 1993. Linear model comparison of black and white dairy bulls for the Nordic countries. *INTERBULL Bulletin* 8.

Boldman, K.G., Freeman, A.E., Harris, B.L. and Luck, A.L. 1992. Prediction of sire transmitting abilities for linear type traits. *Journal of Dairy Science* 75:552-563.

Congleton, W.R. and King, L.W. 1984. Profitability of dairy cow herd life. *Journal of Dairy Science* 67:661-674.

Druet, T. 1998. Prédictions précoces de valeurs génétiques de survie de bovins laitiers. M.S. thesis. Faculté Universitaire des Sciences Agronomiques de Gembloux, Belgium.

Ducrocq, V. 1987. An analysis of length of productive life in dairy cattle. Ph.D.Diss. Cornell Univ., N.Y., USA.

Harris, B., and Johnson, D. 1998. Approximate reliability of genetic evaluations under an animal model. *Journal of Dairy Science* 81:2723-2728.

Jairath, L., Dekkers, J.C.M., Schaeffer, L.R., Liu, Z., Burnside, E.B. and Kolstad, B. 1998. Genetic evaluation for herd life in Canada. *Journal of Dairy Science* 81:550-562.

Klei, B. 1998. Solving MACE equations. *INTERBULL Bulletin* 17:3-11.

Lien, S., Gomez-Raya, L., Steine, T., Fimland, E. and Rogne, S. 1995. Associations between casein haplotypes and milk yield traits. *Journal of Dairy Sciences* 78:2047-2056.

Miesenberger, J., Sölkner, J. and Essl, A. 1998. *INTERBULL Bulletin* 18:78-84

Rogers, G.W., Banos, G., Sander-Nielsen, U. and Philipsson, J. 1996. Genetic correltions among somatic cell score, productive life and type traits from the United States and udder health measures from Denmark and Sweden *INTERBULL Bulletin* 14:34-38

Schaeffer, L.R. 1993. Multi-trait Accross Country Evaluation of Dairy Sires. *INTERBULL Bulletin.* 8

Schaeffer, L.R.. 1994. Multiple country comparison of dairy sires. *Journal of Dairy Science* 77:2671-2678.

Sigurdsson, A. and Banos, G. 1995. Estimation of genetic correlations between countries. *INTERBULL Bulletin* 11

Smith, S.P. 1983. The extension of failure time analysis to problems of animal breeding. Ph.D.Diss. Cornell Univ., N.Y., USA.

VanRaden, P.M. and Wiggans, G.R. 1991. Derivation, calculation, and use of national animal model information. *Journal of Dairy Science* 74:2261-2267.

Vollema, A.R. and Groen, A.F. 1997. Genetic correlations between longevity and conformation traits in an upgrading dairy cattle population. *Journal of Dairy Science* 80:3006-3014.

Weigel, K.A. 1996. Use of correlated trait information to improve the accuracy of early prediction of breeding vale for length of productive life. *INTERBULL Bulletin* 12:125-135.

Weigel, K.A., Lawlor, T.J., VanRaden, P.M. and Wiggans, G.R. 1998. Use of linear type and production data to supplement early predicted transmitting abilities for productive life. *Journal of Dairy Science* 81:2040-2044.