

MACE Weighting Factors for Direct and Maternal EBVs from Multiple-Trait Models

P.G. Sullivan^{1,2}, G.J. Kistemaker¹, J. Jakobsen³ and F. Fikse³

¹Canadian Dairy Network, Guelph, Canada

²Beef Improvement Ontario, Guelph, Canada

³Interbull Centre, Uppsala, Sweden

Introduction

When MACE was first applied, number of daughters (ND) were used as weighting factors for national EBV of sires. Interbull (2000) replaced ND with effective daughter contributions (EDC), which accounted for ND, but also for the number of effective records each daughter contributed to the sire's genetic evaluation. Effective records were adjusted for the estimation of fixed effects, correlated or repeated trait information, performance information of sires' mates, and linear function definitions of index traits. Hence, the EDC accounted for major differences in the national genetic evaluation models of most countries.

Procedure modifications were required, however, to compute EDC for models that included random regressions of genetic effects (Kistemaker and Sullivan, 2000; Strandén *et al.*, 2000). Similarly, modifications will be needed to compute EDC for models with both direct and maternal genetic effects.

The purpose of this paper is to suggest improvements to the EDC calculations for multiple-trait models in general, and to extend the procedures to accommodate models with direct and maternal effects, such as calving ease.

Multiple-trait models

Central to the EDC procedure is the calculation of each animal's reliability based on its own performance ($R(o)$). Selection index equations are used in this step. For multiple-trait models $R(o) = \mathbf{k}'\mathbf{C}'\mathbf{P}^{-1}\mathbf{C}\mathbf{k} / \mathbf{k}'\mathbf{G}\mathbf{k}$. In some but not all cases, $\mathbf{C} = \mathbf{G}$ (e.g. Interbull, 2000). The matrix of covariances among the animal's observations (\mathbf{P}) described by Interbull (2000) correctly accounts for effective observations (m) that are different than 1 on the diagonals,

but not on the off-diagonals. This simplification can lead to negative-definite \mathbf{P} -matrices and negative values for $R(o)$ for animals with repeated records, if residual correlations are non-zero. Examples where m can deviate significantly from 1 are multiple-lactation repeatability models, random regression test-day models, maternal effects models and sire (paternal effects) models.

A better definition of \mathbf{P} for multiple-trait repeatability models is as follows:

$$\text{Let } \mathbf{y}^* = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$$

$$y_{ijk}^* = a_{ij} + pe_{ij} + e_{ijk}$$

where y_{ijk} is the k^{th} observation for trait j and animal i , and a_{ij} , pe_{ij} and e_{ijk} are additive genetic, permanent environmental and residual effects. Covariances among traits for a , pe and e are described by matrices \mathbf{G} , \mathbf{E} and \mathbf{R} respectively. Now define an incidence variable for the j by j' trait combination observed for animal i :

$$v_{ijj'} = \frac{\sum_{k=1}^{n_{jj'}} w_{ikj'}}{\sum_{k=1}^{n_j} w_{ikj} \sum_{k=1}^{n_{j'}} w_{ikj'}}$$

where w_{ikj} is defined by Interbull (2000) and

$$w_{ikj} = \sqrt{w_{ikj} w_{ikj'}}. \text{ Now:}$$

$$P_{ijj'} = G_{ijj'} + E_{ijj'} + v_{ijj'} R_{ijj'}$$

which is equivalent to Interbull (2000) for diagonal elements of \mathbf{P} , but accounts for incidences of trait combinations in the off-diagonals that deviate from 1.

Models with direct and maternal effects

When both direct and maternal effects are of interest, the required matrices for $R(o)$ can be defined as follows:

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_{11} & & \text{sym.} \\ \vdots & \ddots & \\ \mathbf{P}_{41} & \cdots & \mathbf{P}_{44} \end{bmatrix} = \mathbf{V}(\mathbf{y}^*)$$

$$\mathbf{G} = \begin{bmatrix} \mathbf{G}_{11} & \text{sym.} \\ \mathbf{G}_{21} & \mathbf{G}_{22} \end{bmatrix} = \mathbf{V}(\mathbf{g})$$

$$\mathbf{C} \approx \mathbf{C}(\mathbf{y}^*, \mathbf{g}')$$

corresponding with:

- \mathbf{g}_1 = direct genetic effects for n traits
- \mathbf{g}_2 = maternal genetic effects for m traits
- \mathbf{y}_1 = weighted average of observations for direct effects (traits 1 to m)
- \mathbf{y}_2 = weighted average of observations for direct effects (traits $m+1$ to n)
- \mathbf{y}_3 = weighted average for maternal effects on natural progeny (traits 1 to m)
- \mathbf{y}_4 = weighted average for maternal effects on ET progeny (traits 1 to m)

Each trait-by-trait element of sub-matrices in \mathbf{P} is a weighted average of the corresponding elements in \mathbf{G} , \mathbf{E} and \mathbf{R} , using the weights in Table 1. Similarly, matrix \mathbf{C} is derived from elements in \mathbf{G} . For matrix \mathbf{C} , covariances between \mathbf{g} and the parts of \mathbf{y} that are not the animal's own performance are set to zero, to avoid double counting of information in subsequent steps of the EDC procedure.

Table 1. Weights^z to derive trait-by-trait elements of \mathbf{P} and \mathbf{C} matrices.

	\mathbf{G}_d	\mathbf{G}_{dm}	\mathbf{G}_{md}	\mathbf{G}_m	\mathbf{E}_d	\mathbf{E}_m	\mathbf{R}
\mathbf{P}_{11}	1	$\{\frac{1}{2}\}$	$\{\frac{1}{2}\}$	1	1	1	ν
\mathbf{P}_{21}	1	$\{\frac{1}{2}\}$.	.	1	.	ν
\mathbf{P}_{22}	1	.	.	.	1	.	ν
\mathbf{P}_{31}	$\frac{1}{2}$	$\{\frac{1}{4}\}$	1	$\{\frac{1}{2}\}$.	.	.
\mathbf{P}_{32}	$\frac{1}{2}$.	1
\mathbf{P}_{33}	$\frac{1+3\nu}{4}$	$\{\frac{1}{2}\}$	$\{\frac{1}{2}\}$	1	ν	1	ν
\mathbf{P}_{41}	.	.	1	$\{\frac{1}{2}\}$.	.	.
\mathbf{P}_{42}	.	.	1
\mathbf{P}_{43}	.	.	$\{\frac{1}{2}\}$	1	.	.	.
\mathbf{P}_{44}	ν	.	.	1	ν	1	ν
\mathbf{C}_{11}	1
\mathbf{C}_{12}	1
\mathbf{C}_{21}	1
\mathbf{C}_{22}	1
\mathbf{C}_{31}	.	.	1
\mathbf{C}_{32}	.	$\frac{1}{2}$.	1	.	.	.
\mathbf{C}_{41}	.	.	1
\mathbf{C}_{42}	.	.	.	1	.	.	.

^zWeights in $\{\}$ are set to zero (.) for ET animals.

Animal versus progeny performance

In models for traits measured in both sexes, sires can have both individual and progeny performance included in the analysis. Both sources of information contribute to the EDC of the sire. The contribution for progeny i (Interbull, 2000) is:

$$\frac{kR_i(o)}{4 - R_i(o)[1 + R_{dam}(o)]}$$

The additional contribution from the sire's own performance is:

$$\frac{kR_{sire}(o)}{1 - R_{sire}(o)}$$

Sire-MGS models

Calculation of EDC for these models is straightforward. The incidence variables in ν are accumulated with reference to sire and maternal grandsire instead of animal and dam, and animal-model matrices \mathbf{G} and \mathbf{R} are replaced by $\mathbf{S}=\mathbf{LGL}'$ and $(\mathbf{G}-\mathbf{S}+\mathbf{R}+\mathbf{E})$ respectively. For multiple-trait models with direct and maternal effects,

$$\mathbf{L} = \begin{bmatrix} \left(\begin{array}{cc} \frac{1}{2}\mathbf{I}_m & \mathbf{0} \\ \mathbf{0} & \frac{1}{2}\mathbf{I}_{n-m} \end{array} \right) & \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix} \\ \left(\begin{array}{cc} \frac{1}{4}\mathbf{I}_m & \mathbf{0} \end{array} \right) & \frac{1}{2}\mathbf{I}_m \end{bmatrix} \text{ for } \mathbf{G} = \begin{bmatrix} \mathbf{G}_{dd} & \mathbf{G}_{dm} \\ \mathbf{G}_{md} & \mathbf{G}_{mm} \end{bmatrix}$$

Then $R(o)$ is computed the same way as for an animal model, and for each sire:

$$\text{EDC} = \frac{kR_{sire}(o)}{1 - R_{sire}(o)}$$

Threshold models

For these models, the same procedures apply. However, each record is subject to a potentially different residual covariance matrix, based on probability distributions of the categorical observations. The covariance matrices used for EDC calculations should correspond to the scale of expression (underlying or observed) of the national EBV. If the national EBV are expressed on the observed scale, it may suffice to compute EDC as if a linear model had been used since reliabilities of sire EBVs are generally similar for linear relative to threshold models (Phocas and Laloë, 2003; Varona *et al.*, 1999).

Application and testing

A general-purpose program was developed to test the above theory and simplify its application in the field. Three files are required to run the program: a data file, a pedigree file, and a parameter file that specifies the type of national model used (animal or sire), the total number of traits, the number of traits with direct and maternal effects, definitions of any index traits of interest and the covariance parameters used in the national evaluation.

Data were generated to test the program for a variety of alternative models. Six groups of females were created, two each of size 4, 8 and 13, with 1, 2 and 4 of the females being ET recipients. Females survived a maximum of 3 years with a survival probability of 0.8 each year after the first. Each year, and separately for each group of females, one high, one medium and one low usage sire was chosen to mate with the survivors of the group. Sire mating probabilities within the group matched a usage ratio of 3:2:1 for the high, medium and low usage sires, respectively. The sires of the females were subject to the same usage ratio, so many of the females were paternal half sibs, and each female had a different, unrelated dam. Sires and mates of females were sampled from two unrelated groups, so progeny were not inbred. Each of the simulated females and each of their progeny generated a maximum of 4 performance records (one per year), with probabilities of 0.9, 0.7, 0.8 and 0.5. Progeny records began 1 year after the females' first record. Any combination of records was allowed, subject to the above probabilities. Contemporary groups (CG) for genetic evaluation were defined by the interaction of female group and year. Weights of 0.6, 0.8 or 1.0 were assigned, with equal probability, to each record. The idea was to generate variability in CG size, progeny per sire, per dam and per maternal grandsire, sire progeny per CG, records and weights on records per animal and dam, and a combination of natural and embryo transfer data. The simulated data set was only replicated once, since the Monte Carlo steps were only used to create variability in the above parameters.

From the simulated data, several input data sets were created to compute EDCs under a variety of models, and a small subset of these data sets was chosen for discussion in this paper. To assess the merits of weighting factors being considered, "approximate" reliabilities were derived, from inverted sire-(mgs-)model equations based on the weighting factors, and compared with "true" reliabilities, from inverted equations based on performance records. The approximate reliabilities should be highly correlated and have similar means compared with the true reliabilities, and ideally there should be consistent agreement for all types of models. This latter point is critical to

minimize advantages and disadvantages among countries participating in MACE, which could otherwise result from differences in the types of models being used by each country.

Correlations and relative differences in means of approximate and true reliabilities, for sires of cows and mates of cows, are in Table 2. For both sires and mates, correlations were extremely low and average reliabilities severely underestimated for the current weighting factors (EDC2000) when correlations between traits were non-zero (Data **MM) and a sire model was used. Strange results were expected for these situations because incidence information under a sire model is based on progeny counts and therefore deviates significantly from I , the value assumed for incidences of trait combinations in EDC2000. Negative-definite covariance matrices resulted and led to values of $R(o)$ that were outside of the parameter space. The proposed correction in EDC2004 effectively resolved the problem. The corresponding correlations were near unity and average biases close to zero for EDC2004. Differences between EDC2004 and EDC2000 were much smaller under an animal model, where effective records per animal did not deviate significantly from I .

Correlations between approximate and true reliabilities were close to unity and consistently highest for EDC2004, and the variation in bias across all data situations was the lowest for EDC2004. There were biases, however, with reliabilities of sires underestimated and mates overestimated, most notably under an animal model. Underestimation of sire reliabilities under an animal model is explained by the fact that maternal grand-progeny information is ignored when male-only EDC and male-only pedigree information are used to build the mixed-model equations. The overestimation of mate reliabilities seemed to be due to the fact that the female mates were related, since similar biases were not observed in data sets with unrelated females.

The sires were one generation older than the mates, which suggests that the observed pattern of bias could influence trends in sire variances estimated by Interbull. The expected trend in bias would cause a decreasing trend in

sire variance estimates for recent years, for countries that have data structures similar to the simulated data in the present study.

Results show that EDC2004 could be used for multiple-trait models that include both direct and maternal effects, and that it would be superior to ND. A general-purpose program has been written to compute EDCs with the proposed methods, for any multiple-trait animal/dam or sire/mgs model, random-regression models excepted. Plans are to update the program for huge-scale applications so that all countries can use it in the near future.

Conclusions

The general EDC approach outlined by Interbull (2000) has been successfully extended for sire models and for traits that involve both direct and maternal effects. The approach can be used for threshold models. Modifications to the procedure were necessary to prevent major errors under multiple-trait repeatability models, for example a multiple-trait sire model. The procedures were also updated to avoid double counting of information between the various steps when direct and maternal effects are considered simultaneously. The revised procedure allows for the inclusion of embryo transfer data for countries that correctly include those records in their national evaluation systems. The procedure was equally effective for maternal as for direct traits in models containing both effects, and was also useful for an index trait that was the average of all traits in the model. Improvements would be desirable to reduce biases for index traits under sire and animal models, and for all traits under animal models.

International evaluations could be improved with use of the same EDC program by all countries, assuming that such a program was fully tested for the types of national models being applied. A single, general-purpose program will also simplify the expansion of international evaluation services to include new traits of interest, for example fertility traits.

References

- Kistemaker, G.J. & Sullivan, P.G. 2000. Calculation of weighting factors for the Canadian test day model. *Interbull Bulletin* 26, 75-77.
- Interbull, 2000. New weighting factors for the international genetic evaluation; revised July, 2000. *Mimeo*.
- Phocas, F. & Laloë, D. 2003. Evaluation models and genetic parameters for calving difficulty in beef cattle. *J. Anim. Sci.* 81, 933-938.
- Strandén, I., Lidauer, M., Mäntysaari, E.A. & Pösö, J. 2000. Calculation of Interbull weighting factors for the Finnish test day model. *Interbull Bulletin* 26, 78-80.
- Varona, L., Misztal, I. & Bertrand, J.K. 1999. Threshold-linear versus linear-linear analysis of birth weight and calving ease using an animal model: II. comparison of models. 1999. *J. Anim. Sci.* 77, 2003-2007.

Table 2. Squared correlation ($100 \cdot R^2$) between approximate and true reliabilities, and relative difference in means ($100 \cdot (\text{approximate} - \text{true}) / \text{true}$) for three different weighting factors^w under an animal model (AM) or a sire model (SM).

Data ^z Trait ^y		100*R ² (%)						Relative difference in means (%)						True ^x Rel. (%)	
		EDC2004		EDC2000		ND		EDC2004		EDC2000		ND			
		AM	SM	AM	SM	AM	SM	AM	SM	AM	SM	AM	SM		
*** SIRES OF FEMALES (n=12) ***															
10L-	1	90	100	90	100	83	96	-26	0	-26	0	0	36	5 ± 2	
10M-	1	94	100	94	100	89	97	-14	2	-14	2	5	30	18 ± 7	
10H-	1	96	100	96	100	94	97	-4	3	-4	3	8	25	31 ± 9	
20ML	1	94	100	94	100	89	97	-14	2	-14	2	5	30	18 ± 7	
20ML	2	89	100	89	100	87	95	-20	1	-20	1	1	34	15 ± 7	
20ML	ave	91	100	91	100	79	84	-18	2	-18	2	-18	-7	16 ± 6	
20MM	1	95	99	94	0	87	95	-13	0	-11	5	-1	21	20 ± 7	
20MM	2	93	100	93	6	80	91	-14	-1	-12	-35	-15	7	17 ± 7	
20MM	ave	95	99	94	7	73	82	-21	-10	-16	-50	-26	-12	20 ± 7	
42MM	d1	96	100	95	25	88	94	-8	-3	-2	-59	-19	-1	17 ± 6	
42MM	d2	95	100	95	4	76	82	-8	-3	-2	-53	-27	-16	16 ± 6	
42MM	d3	96	100	96	9	94	93	-8	-1	-4	-109	-3	17	19 ± 7	
42MM	d4	96	100	96	19	80	87	-7	-3	-2	-101	-26	-10	17 ± 6	
42MM	m1	98	99	93	17	95	94	5	-5	22	-121	90	27	15 ± 6	
42MM	m2	99	99	93	11	82	77	2	-8	22	-118	118	35	13 ± 5	
42MM	ave	92	100	91	14	13	17	-18	-15	-13	-16	-39	-51	20 ± 7	
*** MATES OF FEMALES (n=10) ***															
10L-	1	100	100	100	100	78	78	04	03	04	03	140	142	5 ± 3	
10M-	1	99	99	99	99	78	78	16	11	16	11	95	104	20 ± 9	
10H-	1	95	98	95	98	77	77	25	16	25	16	66	81	33 ± 12	
20ML	1	99	99	99	99	78	78	16	11	16	11	95	104	20 ± 9	
20ML	2	99	100	99	100	78	78	15	11	15	11	118	129	16 ± 9	
20ML	ave	99	100	99	100	79	76	15	11	15	11	84	72	17 ± 9	
20MM	1	98	99	97	16	80	79	17	10	24	-65	85	94	21 ± 9	
20MM	2	98	99	98	88	78	77	19	08	27	-47	82	91	19 ± 9	
20MM	ave	98	99	96	48	78	76	10	-02	22	-18	60	63	21 ± 10	
42MM	d1	95	98	92	2	74	78	31	06	47	-132	74	79	18 ± 8	
42MM	d2	95	98	92	2	69	74	34	7	52	-86	77	70	17 ± 8	
42MM	d3	93	98	88	2	56	65	38	12	54	-77	103	96	18 ± 8	
42MM	d4	93	98	90	0	79	85	37	9	53	-72	61	61	16 ± 8	

^zFour codes for: numbers of direct and maternal effect traits, levels of heritabilities and correlations. Levels of heritabilities for total genetic effects (for direct traits, direct+maternal traits) are: L(.05,.08), M(.25,.30), H(.50,.60).

Levels of correlations for (residuals, direct and maternal effects, direct X maternal) are: L(.0,.0,.0), M(.4,.7,.1).

^yTrait prefixes 'd' and 'm' signify direct and maternal effects. The trait 'ave' is a simple average of all traits.

^xValues reported are for the animal model analysis.

^wEDC2004 are the proposed weighting factors, EDC2000 are the Interbull (2000) procedures extended to allow for maternal effects but without the proposed corrections for multiple-trait models, and ND are numbers of daughters.