Weighting Factors in International Genetic Evaluations: Effects on International Breeding Value and Reliability Estimates

Freddy Fikse and Georgios Banos
Interbull Centre, Sweden

Introduction

National genetic evaluation results can be combined in a multi trait sire model (MACE) to compute international genetic evaluations (Schaeffer, 1994). Variability in accuracy of national breeding values can be accommodated in the variance structure of the residuals in the international sire model. At present, total number of daughters is being used in routine evaluation done by INTERBULL, but other weighting factors are under investigation (Fikse and Banos, 1997; Sullivan *et al.*, 1998; Fikse and Banos, 1999).

Fikse and Banos (1999) studied several alternative weighting factors and their effect on genetic parameters. They found that estimates of genetic correlations did not change, however, genetic variances were affected. The presently used weighting factor, total number of daughters, resulted in overestimated sire variances, whereas total number of lactations gave underestimated sire variances. Weighting factors that consider contemporary group structure, correlation between records and reliability of mate of bull performed best and yielded unbiased estimates of sire variances.

In this study we investigated the impact of weighting factors on international breeding value and reliability estimates.

Material and Methods

Simulation

A conventional progeny test scheme was simulated in two populations that were of equal size. The simulation comprised ten generations, with 36,000 cows and 300 progeny test bulls per generation in each population. Every generation 10 sires and 150 dams were selected as parents of young bulls in the next generation. In addition, 30 bulls were chosen based on progeny test results to be used as proven bulls in the next generation. Progeny group size for young bulls was normally distributed with mean 80 and SD 5. Proven bulls had 400 additional progenies in the following generation.

Selection within population was on predicted breeding values, and was carried out in each generation. Mating was random, and resulted in one offspring for the generation of females and two offspring for males. From generation four onwards, sires of sons and sires of cows were exchanged between population at a fixed rate of 50% to ensure sufficient genetic links between populations.

True breeding values were simulated as: $TBV_{anim} = .5\ TBV_{sire} + .5\ TBV_{dam} + \Phi_{anim}$, where TBV_{sire} and TBV_{dam} are the true breeding value for sire and dam, and Φ_{anim} represents the Mendelian

sampling term. This term was sampled from a pseudo normal distribution with mean zero and SD $.5 * \sigma_a * \sqrt{(1-(F_{sire}+F_{dam})/2)}$, where σ_a is the genetic SD, and F_{sire} and F_{dam} is the inbreeding coefficient of the sire and dam. Phenotypic variance and heritability were assumed to be 100 and .30, respectively. The genetic correlation between populations was .90.

Phenotypic records for cows were generated as: $y = CG + TBV_{anim} + error$, with CG the contemporary group effect. For each cow three lactation records were simulated as repeated observations with r = .50. Contemporary group size was normally distributed with mean 10 or 20 and CV 5%, with a minimum size of five. Contemporary group effect was generated through a normally distributed random deviate with mean and SD equal to zero and 15, respectively.

At the end of every generation, breeding values were predicted based on first lactation records only. The model included fixed contemporary groups and genetic groups for missing parents defined according to generation and population, and a random animal effect. Simulated genetic and residual variances were used. Fifteen replicates of each scenario were carried out.

National breeding value estimation

Following the simulation, three alternative breeding value estimations within each population were performed. First, records on first lactation only were used, and the model included fixed contemporary group and phantom parent group effects, as well as a random animal effect. Second, records on all three lactations were used, applying the same model as above, including an additional permanent environmental effect. In the last alternative three lactations were considered, but culling on phenotypic records was practiced prior to the breeding value estimation. The

culling process was simulated by deleting certain portions of the second and third lactation records; 50% of the cows were allowed to have records on second lactation and 30% on the third.

Weighting factors:

Seven different weighing factors were considered:
1. Total number of daughters;
2. Total number of lactations;
3. Number of first lactations, adjusted for contemporary group size, and computed as:

$$n_{3_i} = n_i - \Sigma_k \frac{1}{n_{jk}}$$, where the summation is

over daughter k of sire i, and n_{jk} is the size of the contemporary group in which daughter k made her first lactation.

4. Number of first lactations, adjusted for contemporary group size and distribution of daughters over contemporary groups, computed

as: $$n_{4_i} = n_i - \Sigma_j \frac{n_{ij}^2}{n_j}$$, where n_i is the total

number of daughters of sire i, and the summation is over contemporary group j in which sire i has n_{ij} daughters, and n_j is the contemporary group size.

5. Effective number of all lactations, computed

as: $$n_{5_i} = \Sigma_k \left[\frac{k_d REL_{k_{yield}}}{4 - 2 REL_{k_{yield}}} \right]$$, where

$$REL_{k_{yield}} = \frac{n_k h^2}{1 + (n_k - 1)r}$$, with n_k the

number of lactations for daughter k of sire i adjusted for sizes of contemporary groups, $k_d = (4 - 2h^2)/h^2$, and r the repeatability.

6. $$n_{6_i} = \Sigma_k \left[\frac{k_d REL_{k_{yield+yield,dam}}}{4 - 2 REL_{k_{yield+yield,dam}}} \right]$$, where

$$REL_{k_{yield+yield,dam}} = \frac{DE_{k_{yield+yield,dam}}}{DE_{k_{yield+yield,dam}} + k_d} \quad , \text{and}$$

$$DE_{k_{yield+yield,dam}} = k_d \times \left(\frac{REL_{k_{yield}}}{1 - REL_{k_{yield}}} + \frac{REL_{dam_{yield}}}{4 - REL_{dam_{yield}}} \right)$$

7. As in 6., but REL_{dam} instead of $REL_{dam_{yield}}$,

that is, REL_{dam} is not based on the yield of the dam, but also of the grandam, great grandam, etc.

International genetic evaluation

International breeding values were calculated with the Mace model (Schaeffer, 1994), applying iteration on data. De-regressed national breeding values and estimated sire variances were used as in Fikse and Banos (1999). The true genetic correlation was used in the Mace analysis. International reliabilities were computed with the approximation method described by Harris and Johnson (1998).

Genetic groups in the de-regression, sire variance estimation and Mace analysis were defined according to population of origin, generation and selection path.

Results and Discussion

Table 1. Prediction error variances of international breeding values on foreign scale for bulls with evaluation in only one population (mean and SD of 15 replicates); Mace analysis based multiple lactation national breeding values

Weighting factor	local scale				foreign scale			
	CG size 10		CG size 20		CG size 10		CG size 20	
n_{dau}	8.1	.4	7.7	.3	24.2	1.1	23.8	1.0
n_{lact}	8.1	.4	7.7	.3	24.1	1.1	23.8	1.0
DE_{yield}	8.1	.4	7.7	.3	24.1	1.1	23.8	1.0
$DE_{yield} + DE_{yield,dam}$	8.1	.4	7.7	.3	24.1	1.1	23.8	1.0
$DE_{yield} + DE_{dam}$	8.1	.4	7.7	.3	24.1	1.1	23.8	1.0

For bulls with information in only one country, PEV on both the local and foreign scale was unaffected (Table 1). Both the PEV and its standard deviation over replicates was smaller for the alternative with larger contemporary group sizes, as expected. Also for the Mace analyses with national breeding values based single lactation and multiple lactation after culling showed no difference in PEV for the various weighting factors. PEVs for bulls with an evaluation in both countries were lower than PEVs on local scale for bulls with only one

evaluation.

Biases in international breeding values were not affected by choice of weighting factor, and were zero in most cases (results not shown). The slope of the conversion equations is nearly unaffected by choice of weighting factor (Table 2). The theoretical expectation for this slope is $r_G \times \sigma_{Imp}/\sigma_{Exp}$, and because the same weighting factor was always used in both populations, the regression coefficient is not affected. This also explains why only minor changes were observed for international breeding value, PEVs (Table 1), biases, and top 100 ranking.

Table 2. Slope of conversion equations[a] (mean of 15 replicates, SD ≈ .04); Mace analysis based multiple lactation national breeding values

Weighting factor	$CBV_{Pop2} =$ $f(EBV_{Pop1})$		$CBV_{Pop1} =$ $f(EBV_{Pop2})$	
	10[b]	20	10	20
n_{dau}	.84	.86	.87	.84
n_{lact}	.84	.86	.86	.84
DE_{yield}	.84	.86	.86	.84
$DE_{yield} + DE_{yield,dam}$	.84	.86	.86	.84
$DE_{yield} + DE_{dam}$	.84	.86	.86	.84

[a] Estimated from the last generation of young bulls
[b] Average contemporary group size

Another Mace analysis was done in addition to the previously described ones. Input data to this analysis were national breeding values based on all three lactations for the case with average contemporary group size 20. For population 2, all five weighting factors were considered, but for population 1 always $DE_{yield} + DE_{dam}$ was used.

Table 3. Conversion equations[a] (mean of 15 replicates); Mace analysis based multiple lactation national breeding values, and weighting factor $DE_{yield} + DE_{dam}$ used in population 1

Weighting factor in population 2	$CBV_{Pop2} =$ $f(EBV_{Pop1})$		$CBV_{Pop1} =$ $f(EBV_{Pop2})$	
	a	b	a	b
n_{dau}	7.6	.90	13.0	.81
n_{lact}	11.7	.84	9.5	.86
DE_{yield}	9.9	.87	11.1	.84
$DE_{yield} + DE_{yield,dam}$	10.2	.86	10.8	.84
$DE_{yield} + DE_{dam}$	10.2	.86	10.8	.84

[a] Estimated from the last generation of young bulls

When different weighting factors were used in both populations, the regression coefficient changed depending on weighting factor (Table 3). For n_{dau}, the sire variance in population 2 was overestimated (Fikse and Banos, 1999), and the slope of the regression of CBV_{Pop2} on EBV_{Pop1} ($b_{2|1}$) was underestimated. The sire variance for weighting factor n_{lact} was underestimated (Fikse and Banos, 1999), and the regression coefficient was overestimated. Higher slopes went together with lower intercepts, which conforms to the usual tendency that when slopes change, intercepts change in the opposite direction. Converted breeding values for top bulls were therefor nearly the same independent from weighting factor.

Table 4. Number of local bulls in top 10/100 on local scale[a]; Mace analysis based multiple lactation national breeding values, and weighting factor DE_{yield} + DE_{dam} used in population 1

Weighting factor	Pop. 1		Pop. 2	
	10[b]	100	10	100
n_{dau}	7.1	64	7.3	61
n_{lact}	6.4	60	7.6	65
DE_{yield}	6.7	61	7.5	63
$DE_{yield} + DE_{yield,dam}$	6.6	61	7.6	64
$DE_{yield} + DE_{dam}$	6.6	61	7.6	64

[a] Top bull lists considered both proven and young bulls
[b] Size of top bull list

It appeared that population 1 was slightly inferior to population 2, although this was not intended in the simulation. This explains why bulls from population were more frequently represented on the local scale (Table 4). Because the regression coefficient $b_{2|1}$ was underestimated, the variance of foreign-scale breeding values was smaller, resulting in fewer bulls from population 2 on both rankings when n_{dau} was used as weighting factor. The reverse is true for n_{lact}, but for the other three weighting factors the observed values were closer to the expected.

Table 5. International reliabilities on local and foreign scale[a] for bulls with information in only one population; Mace analysis based multiple lactation national breeding values

Weighting factor	local scale		foreign scale	
	10[b]	20	10	20
n_{dau}	87	88	74	74
n_{lact}	95	95	79	79
DE_{yield}	92	92	77	77
$DE_{yield} + DE_{yield,dam}$	93	93	77	77
$DE_{yield} + DE_{dam}$	93	93	77	77
Expected	92	92	76	76

[a] Mean over 15 replicates; SD < .05
[b] Average contemporary group size

Reliabilities on both local and foreign scale were affected by choice of weighting factor. n_{dau} and n_{lact} under- respectively overestimate the domestic reliability, and the same is found for reliabilities of international breeding values compared when compared to the expectation.

The expected value for the reliability is based on the observed PEV for the same group of bulls (Table 1), using the relationship: REL = (σ^2 - PEV) / σ^2. International reliability and PEV of international breeding values match better for the other weighting factors. Although less pronounced, the same pattern of international reliabilities and their agreement with the empirical PEV were found for the other national evaluation alternatives.

Apart from total number of daughters, none the considered weighting factors are currently submitted to INTERBULL, and countries would need to provide those. Once the number of contemporaries is available for each lactation record, DE_{yield} can be calculated by processing

the file with lactations. For the weighting factors $DE_{yield} + DE_{yield,dam}$ and $De_{yield} + DE_{dam}$, DE contributed by the dam needs to be added for each cow. Then DE contributions can be collected for bulls. If data files are sorted in a convenient order, the considered weighting can be computed in three passes.

In conclusion, when the same weighting factor was used in both populations, only minor changes in international breeding values were observed. Breeding values were more affected when different weighting factors were used in both populations. International reliabilities were closer to their expectation when contemporary group structure, correlation between records and reliabilities of mates were considered. For n_{dau} and n_{lact}, international reliabilities were off by up to five percent points.

References

Fikse, W.F. & Banos, G. 1997. Alternative weighing factors for daughter information in international genetic evaluations: progress report. Interbull bulletin, 16, 3-7.

Fikse, W.F. & Banos, G. 1999. Weighting factors of daughter information in international genetic evaluation for milk production traits: effect on (co) variance components. J. Dairy Sci. 82 (suppl. 1): 72. (Abstr.)

Harris, B. & Johnson, D. 1998[a]. Information source reliability method applied to MACE. Interbull Bulletin 17, 31-36.

Schaeffer, L.R. 1994. Multiple-country comparison of dairy sires. J. Dairy Sci. 77, 2671-2678.

Sullivan, P.G., Harris, B. & Fikse, W.F. 1998. Aligning domestic and MACE reliabilities for conformation. Paper presented at INTERBULL Workshop Oct. 12-13, Cremona, Italy.

VanRaden, P.M. & Wiggans, G.R. 1991. Derivation, calculation, and use of national animal model information. J. Dairy Sci. 74, 2737-2746.