Alternative Strategies for Estimation of Country Sire Variance in International Evaluations of Dairy Bulls

Filippo Miglior[1], Peter G. Sullivan[2] and Brian J. Van Doormaal[1]
[1]Canadian Dairy Network, Guelph, Ontario, Canada
[2]CGIL, University of Guelph, Guelph, Ontario, Canada

1. Introduction

The MACE program is the most accurate method currently available for international genetic comparisons. The multiple-country model allows each country to have different genetic parameters, different units of measurement and genetic correlations between countries that are less than one. The implementation and further evaluation of MACE has however revealed some practical limitations, such as its sensitivity to variance component estimates.

Sire variance estimates are used in MACE to account for different scales of proof expression in each country. Currently these variances are estimated by REML using all bulls that have daughters in at least 10 herds, by country. This group includes locally sampled bulls and imported bulls first sampled abroad. Sire variances are estimated at each run using approximate Mendelian sampling terms (MS), which have an assumed expected value of 0 for all bulls. While average MS of bulls sampled locally may be close to zero, MS of imported bulls tends to be positive, as only top proven bulls are imported in any country.

Previous research in Canada (3,4) and Italy (1) has shown that various countries have heterogeneous trends in sire variance estimates over time. In addition, the impact of each country's relative sire variance estimate is known to have an important impact on international sire rankings computed by Interbull using MACE.

The objective of this study was to identify an alternative strategy to estimate sire variances in Interbull evaluations in order to decrease the effect of heterogeneous sire variance trends over time and across countries and to diminish the impact of second country proofs on sire variance estimation. A field and a simulation study were carried out.

2. Field study – Materials and Methods

National Holstein bull proofs were kindly provided from ten major dairy countries (Australia, Canada, Denmark, France, Germany, Great Britain, Italy, The Netherlands, New Zealand and United States). A total of 59,657 Holstein bulls were evaluated with MACE based on genetic evaluation files and genetic correlation estimates used by Interbull in the November 2000 routine run, and alternative estimates of sire variances.

Sire variances were estimated for each country using MS of either: a) all bulls, the current Interbull procedure (**ALL**); b) bulls sampled locally (**Local**); or c) bulls born in the last five years (**5 years**). The last group included mostly first crop locally sampled bulls. Variances were estimated within and across years for each of milk, fat and protein yield. The sire variance estimates used were the only difference between MACE analyses. Results from the MACE analyses for each trait were compared based on sire variance trends, relative estimates of sire variance, and differences in country representation in top 100 lists.

3. Field study – Results and Discussion

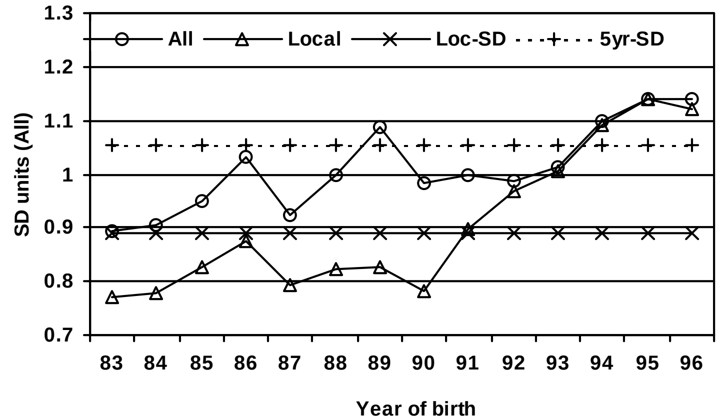
Numbers of bulls by birth year are shown in Table 1. Bulls from these ten countries represented 91% of the total number of bulls from 27 countries

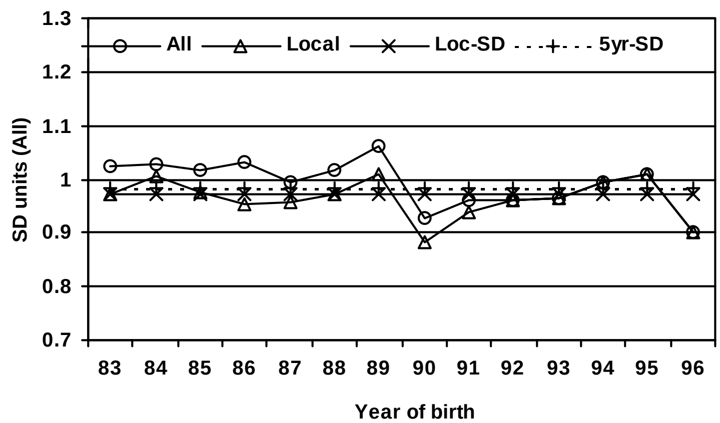
included in the official Interbull run. The group of bulls born in 1996 was far from being complete due to the different times required for proving bulls within and across countries.

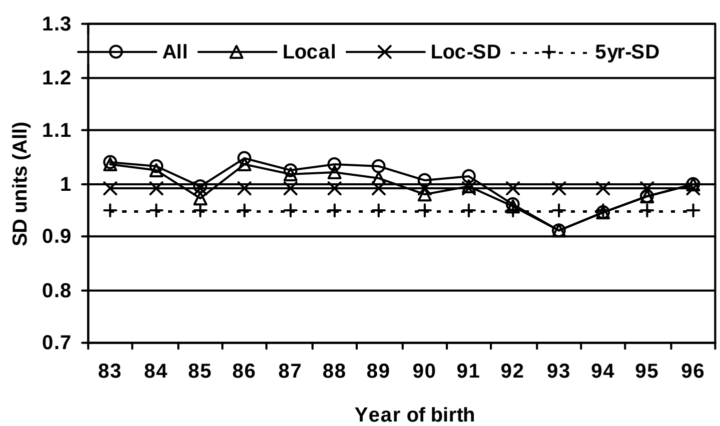
Only results for protein yield are presented, as results for milk and fat yield were very similar. In figures 1, 2 and 3 sire variance trends over time are shown for three countries with an increasing, flat and decreasing trend, respectively. Values have been standardized based on the overall value of sire SD estimated using all bulls (**All**). Thus the value of 1 represents the estimation of sire SD as currently computed by Interbull.

In general, trends of sire SD for local bulls were always lower than the trends for all bulls. This result was expected because the inclusion of MS terms of imported bulls, which are affected by selection, causes an upward bias in sire variance estimates. Estimates of sire variance using only the last 5 years should be closest to the variability of sire expression that is expected for the next generation of matings.

In Table 2 changes in sire variance and average top 100 rankings are shown relative to the current Interbull (all bulls) procedure. Sire variances estimated using locally sampled bulls ranged from -11% to 0% lower than variance estimated using all bulls. Sire variances estimated using bulls born in the last 5 years ranged from -5% to +9% of the variance estimated using all bulls. A 1% decrease in SD produced an average 5% increase in top 100 representation, as estimated by linear regression.







Figures 1, 2 and 3. Sire standard deviation trends for three countries, relative to the overall sire SD when all bulls were used for estimation. Loc-SD is the overall sire SD when only locally sampled bulls were used for estimation, 5yr-SD is the overall sire SD when only bulls born in the last 5 years were used for estimation.

Table 1. Distribution of bulls by birth year and sire variance estimation analysis (**All**, **Local** and **5 years**)

Birth year	**All**	**Local**	
		N	%
1983	2825	2621	93%
1984	3134	2974	95%
1985	3673	3463	94%
1986	4449	4138	93%
1987	4363	4182	96%
1988	4737	4487	95%
1989	4892	4672	96%
1990	5181	4928	95%
1991	5245	5083	97%
1992	**5346**	5328	100%
1993	**5202**	5198	100%
1994	**5303**	5301	100%
1995	**5184**	5184	100%
1996	**1378**	1378	100%

*Note: Bulls in the shaded area are those included in the **5 years** sire variance estimation analysis*

Table 2. Sire SD ratio vs. Top 100 ratio by country scale

	National / All ratio		**5 years / All ratio**	
Country	**Sire SD**	**Top 100**	**Sire SD**	**Top 100**
A	0.89	1.42	1.05	0.58
B	0.93	1.67	0.97	1.28
C	0.96	1.07	1.00	0.87
D	0.95	1.20	0.97	1.15
E	0.97	1.00	1.09	0.81
F	0.96	1.21	1.06	0.59
G	0.97	1.05	0.98	0.93
H	0.98	1.00	1.01	0.87
I	0.99	0.85	0.95	1.21
J	1.00	0.90	1.00	0.92

4. Simulation study - Materials and Methods

The purpose was to compare alternative MACE approaches, assuming there is true heterogeneity of variance in sire expression over time in national evaluations. Variance heterogeneity may exist because of trends in environment, management or recording practices that affect variability of genetic expression, or due to incorrect or missing variance heterogeneity adjustments in national evaluation models.

The national estimated transmitting abilities (ETA) for protein yield from the field study were used to simulate true transmitting abilities (TA) on ten country scales, over 20 replicates. The simulated TA were adjusted to standardize sire variances over time on each country scale, to the prior assumed values that were estimated across all years from the field study.

The TA were sampled from the following distributions: For observed ETA,

$$\mathbf{TA} \sim \mathbf{MVN}(\mathbf{ETA}, \mathbf{DSD})$$

$$\mathbf{D} = \mathrm{diag}\{\sqrt{(1-rel_i)}\}$$

where **S** is the sire covariance matrix and rel_i is the reliability of ETA on country scale i.

Where ETA were not observed,

$$\mathbf{TA} \sim \mathbf{MVN}([\mathbf{PA_m} + \mathbf{S_{mo}S_{00}^{-1}M_o}],[\mathbf{S_{mm}} - \mathbf{S_{mo}S_{oo}^{-1}S_{om}}]\delta)$$

where subscripts m and o represent vectors of missing and observed scales, **PA** is a vector of parent averages and **M** a vector of Mendelian Sampling terms, under a sire model, and δ is a constant ranging from 11/16 to 1 that defines the extent of known immediate ancestry. Implicitly it was assumed that the **ETA** on each scale were unbiased, and that **S** was the matrix of true sire variances and covariances.

The **TA** were subsequently adjusted to remove heterogeneity of sire variance across years, to create $\mathbf{TA_{sd}}$. Sire variances were computed for a given year with the REML equation of Sullivan (2), using the simulated true Mendelian Sampling terms of all local bulls and setting the trace term equal to zero. Bulls were processed one year at a time in chronological order. Elements of $\mathbf{M_o}$ for each bull were multiplied by the corresponding ratios of sire standard deviations across versus within year to standardize the sire variance. It was implicitly assumed by this adjustment that variance of sire expression on each country scale was the same for local bulls as for locally proven foreign bulls. The $\mathbf{TA_{sd}}$, where **ETA** were not observed, were computed as before but using the variance-standardized $\mathbf{M_o}$.

5. Simulation study - Results and Discussion

Sire variances and correlations from the simulated **TA_{sd}** of all local bulls were validated to match the prior across-year values, by and across each of the country scales. The scaling factors that were required to remove heterogeneity of sire variances across years were as expected given estimates of sire variance by year, from the Interbull programs for the current data.

Effects of standardizing variances are largest at extremes of distributions, and in the present study were expected to have the greatest impact on relative country rankings at the extremes. Observations were therefore focused on prediction errors and country representation in top 100 lists of bulls. Differences in prediction error (RMSE) among the alternative MACE evaluations, relative to the simulated TA were small (Table 3). However, expected ranking relative to TA was clearly the closest for M-5yr, based on correlations between the MACE evaluations and the average TA across replicates (Figure 4). This suggests that while all three approaches have similar accuracy, M-5yr results match best with what is expected from a simple blending of national evaluations from multiple countries. The M-5yr results would likely appeal to national clients as the most reasonable set of international evaluations.

Differences in RMSE between MACE evaluations and TA_{sd} clearly favoured M-5yr over the other two MACE alternatives (Table 3). If sire variance heterogeneity in national evaluations is real, M-5yr offers a simple method to account for it in international evaluation.

The impact of using M-5yr instead of M-All was also studied by correlating the changes in country representation in top 100 against the corresponding changes for TA_{sd} versus TA. With prior variances in the simulation from M-All, M-Local and M-5yr, these correlations were respectively .818, .818 and .859, further supporting the conclusion that M-5yr offers an effective way to account for trends in sire variance that may differ between countries. If, on the other hand, sire variance heterogeneity is not real, then M-5yr still provides international evaluations of comparable accuracy (Table 3 and Figure 5).

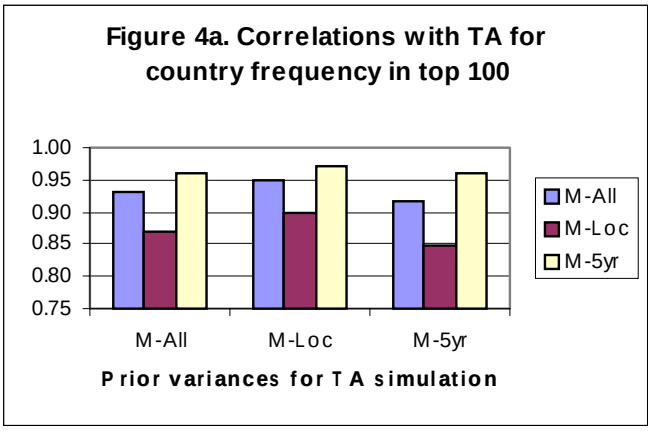
It is of interest to note that for each of the three MACE analyses, approximately 95% of the bulls in the top 100 were born in the last 5 years. It therefore seems logical that the conversion parameters most appropriate to bulls born in the last 5 years would be the best parameters to use in MACE for correct ranking of bulls in the top 100 list, as was observed.

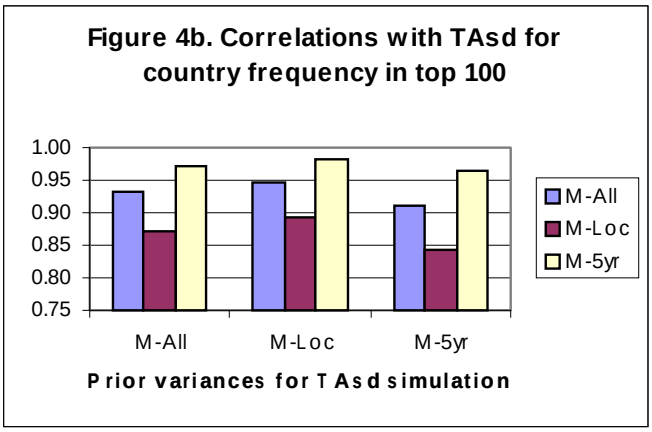
Another benefit of using M-5yr is that it eliminates the need for time-editing data in the international evaluation (2).

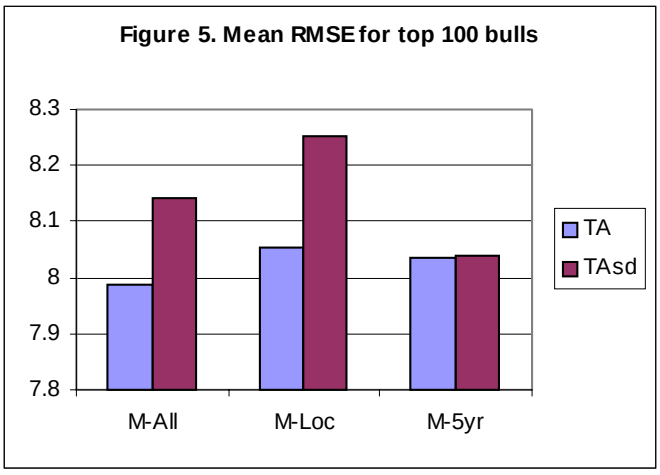
Table 3. RMSE between MACE evaluations and simulated transmitting abilities for the top 100 bulls[z].

Prior Sire Variances	TA versus			TA_{sd} versus		
	M-All	M-Local	M-5yr	M-All	M-Local	M-5yr
M-All	8.02	8.08	8.06	8.18	8.29	8.08
M-Local	7.96	8.04	8.00	8.11	8.23	7.99
M-5yr	7.98	8.04	8.04	8.13	8.23	8.04
Mean RMSE	7.99	8.05	8.03	8.14	8.25	8.04

[z]Results are averages across two sets of top 100 lists, selecting by either M-All or M-5yr evaluations, averaged across country scales with weighting by the inverse of the sire SD on each scale.







6. Conclusions

Estimation of sire variance is critical in international genetic evaluations. The current estimation method used by Interbull overestimates sire variances due to the inclusion of 2nd country proofs of selected bulls, and does not account properly for heterogeneity of sire variances over time and across countries. The latter concern is especially important for the most recent group of bulls that are at the top of international rankings. Changes in relative estimates of sire variance has a significant impact on international rankings of countries, with an increase of 5% of bulls in the top 100 for every 1% decrease of estimated sire SD. Biases in sire variance estimates can be removed and variance heterogeneity effectively accounted for by estimating sire variances for each country using only bulls born in the most recent 5 years.

References

1. Cassandro, M., Miglior, F., Carnier, P., Bittante, G., Canavesi, F., Santus, E. & Banos, G. 1997. Effect of standardization of within country-year sire variance of deregressed proofs on international evaluations. *Interbull Bulletin 16,* 16.

2. Sullivan, P.G. 1999. REML estimation of heterogeneous sire (co)variances for MACE. *Interbull Bulletin 22,* 146.

3. Van Doormaal, B.J., Kistemaker, G.J. & Sullivan, P. G. 1999 Heterogeneous variances of Canadian bull EBV over time. *Interbull Bulletin 22,* 141.

4. Van Doormaal, B. J. & Miglior, F. 2000. Trends in sire variance estimates by birth year. *Interbull Bulletin 25,* 70.