Sire-Dam Pedigree in MACE
– Results from a Full-Scale Pilot Study

Jette Jakobsen and Freddy Fikse
Interbull Centre, Department of Animal Breeding and Genetics, SLU, S-750 07 Uppsala, Sweden

1. Introduction

Earlier studies (e.g. De Jong, 2003) have indicated that international genetic evaluations are sensitive to definition of genetic groups. De Jong (2003) suggested using sire-dam pedigree instead of sire-MGS pedigree in order to get the phantom parent groups further away from the animals with data. Based on this hypothesis Van der Linde *et al.* (2005) collected sire-dam pedigree, made appropriate changes to the HA-USA MACE software (Klei, 1998; Klei & Weigel, 1998) and conducted a pilot study for seven countries participating in the evaluation for the Holstein breed for protein yield. Van der Linde *et al.* (2005) estimated genetic correlations and predicted breeding values using both the sire-dam pedigree and the sire-MGS pedigree. The genetic correlations were rather similar independent of the pedigree structure (average decrease of 0.017) while the CPU time increased with a factor 28. For breeding value prediction the computing time changed with a factor 9 when changing the pedigree structure. Van der Linde *et al.* (2005) also looked at predictability of proofs and found an increase in predictive ability when using the sire-dam pedigree compared to the sire-MGS pedigree.

Based on these findings, the Interbull Technical Committee (ITC) recommended implementation of sire-dam pedigree for prediction of breeding values, but not for estimation of genetic correlations.

Initially the aims of this study were to collect sire-dam pedigree for all bulls currently participating in international genetic evaluation, to study performance for high and low heritability traits and, in order to check feasibility with respect to hard ware and soft ware, to extend the size of the study to include the 25 countries that were participating in the Interbull test run in March 2007. During the study the aims evolved to include the need of improving the software, but also to look at different depth of pedigree and connectedness. Furthermore, the change in reliability was also investigated to quantify the improvement when changing pedigree structure.

2. Material and Methods

2.1 Pedigree Data

All national evaluation centers participating in Interbull Evaluations were asked to submit sire-dam pedigrees in February 2007. In the request for pedigree data we asked to receive at least three generations of pedigree for each bull dam. All but 5 countries supplied bull dam pedigree, and the country specific pedigree files were merged to form one joint international pedigree with sire-dam relationships. The complete pedigree file consisted of just above 530,000 records, of which nearly 270,000 were males.

Currently, national evaluation centers submit birth years on bulls only and the biggest challenge was to obtain consistent birth years also on dams. It was not unexpected to find inconsistencies after merging data from several data providers, however, there were surprisingly many inconsistent birth years within the data submitted by single data providers.

Upon inspection there were a number of birthdates with a high frequency, e.g. January 1st, 1985, and these were assumed to be (incorrectly) estimated and removed; this concerned nearly 17,000 animals.

Subsequently, remaining conflicting birth years (n=850) were identified and solved by the following algorithm:

```
Repeat
  i. Determine minimum and maximum birth
     year based on parents and progeny.
```

ii. Conflict if progeny is born before parents. Solve by either
    a. Deleting birth year of parent or progeny, or
    b. Deleting parent from pedigree
iii. Fill in missing birth years based on parents birth year

The rules for solving the conflict are based on quality and source of information, and also consider whether a 'simple' measure like deleting a birth year actually solves the conflict.

Some of the birth years reported by countries were estimates and not actual known birth years. Including an indicator in the file format for data exchange would alleviate resolving birth year conflicts.

2.2 Data

In order for data to correspond with the pedigree file, data used for the test run in March 2007 was used for this study. Ten of the 34 traits Interbull offered international genetic evaluation for in 2007 were considered in the current study. These were: milk yield, fat yield, protein yield, somatic cell count, mastitis, longevity, direct calving ease, maternal calving ease, direct stillbirth and maternal stillbirth. These traits were chosen as they currently follow a similar workflow for routine evaluation – the only difference being manual vs. automatic forming of phantom parent groups for production and udder health traits vs longevity and calving traits. The number of countries participating as well as the range of heritabilities for the ten traits for Holstein and Red Dairy Cattle are in Table 1.

Table 1. Heritabilities for milk, fat, protein, somatic cells, mastitis, longevity, direct calving ease, maternal calving ease, direct stillbirth, and maternal stillbirth for Holstein (HOL) and Red Dairy Cattle (RDC).

Trait	Range of heritability	No of countries (HOL)	No of countries (RDC)
Milk Yield	0.18 – 0.59	25	11
Fat Yield	0.11 – 0.58	25	11
Protein Yield	0.14 – 0.55	25	11
Somatic Cell Count	0.06 – 0.35	23	10
Mastitis[#]	0.04 – 0.35	23	10
Longevity	0.03 – 0.17	19	9
Direct Calving Ease	0.03 – 0.13	10	3
Maternal Calving Ease	0.02 – 0.12	9	3
Direct Stillbirth[$]	0.02 – 0.11	10	-
Maternal Stillbirth[&]	0.02 – 0.12	9	-

[#] Only DFS participated with mastitis data for the March test run 2007. Other countries submitted proofs for Somatic Cell as best predictor for mastitis. Heritabilities for both direct trait and predictor trait are included in the range.

[$] Countries without information on direct stillbirth can submit direct calving ease as best predictor for direct stillbirth. Heritabilities for both direct trait and predictor trait are included in the range.

[&] Countries without information on maternal stillbirth can submit maternal calving ease as best predictor for maternal stillbirth. Heritabilities for both direct trait and predictor trait are included in the range.

2.3 Scenarios

Genetic correlations for all scenarios were estimated using the procedure as used today with a sire-MGS pedigree and tracing pedigree from 1970 onwards, as per ITC recommendation.

Four different scenarios were performed for breeding value prediction:

1. Sire-MGS pedigree, tracing two generations of pedigree and use of the ITB software
2. Sire-MGS pedigree, tracing the pedigree as far back as possible and use of the HA-USA software
3. Sire-dam pedigree, tracing the pedigree as far back as possible and

use of the HA-USA software adapted for sire-dam pedigree

4. Sire-dam pedigree, tracing the pedigree from 1970 onwards (as is currently done for correlation estimation) and use of the HA-USA software adapted for sire-dam pedigree

2.4 Methods

The software in scenario three and four were modified to be able to handle the change from sire-MGS pedigree to sire-dam pedigree (Van der Linde *et al.,* 2005). This software was further modified to optimize for speed:

- The de-regression is based on the original description by Sigurdsson and Banos (1995), but modified to include the mean. Thus, given the current estimate for the mean, de-regressed national evaluations are computed in "one-shot" (i.e. not iteratively). The de-regressed national evaluations are subsequently used to obtain a new estimate for the mean. These two steps (computing de-regressed breeding values given μ and updating μ) are repeated until the estimate of the mean has converged.
- Sire variances were estimated within-country using a one-step procedure (Sullivan, 1999). This approach exploits the situation that heritability and breeding values are known (i.e. reported by each country), thereby avoiding the need for iterative updating. An appealing by-product is within-year estimates of sire variances.
- Strategies for solving MACE mixed model equations as outlined by Fikse (2008)

Phantom parent groups were formed as breed of bull x birth year of bull x country of origin of maternal grand dam and automatically assigned in scenario two, three and four and of the same size as used for correlation estimation. These are currently of size 15, 30, 30, 30 for production, udder health, longevity and calving traits, respectively.

Approximate reliabilities of MACE breeding values were computed using the information source method outlined by Harris and Johnson (1998). The computations were adapted to sire-dam pedigree by Van der Linde *et al.* (2005). Reliabilities were computed along with the breeding value prediction for all four scenarios of software and pedigree.

3. Results and Discussion

3.1 Change from ITB software to HA-USA software

For both scenario one and scenario two a sire-mgs pedigree was used. The main differences between both were the depth of pedigree used as well as the software package. Product moment correlations were computed across all bulls on all country scales for Holstein for protein yield and were in the range from 0.978 to 0.995.

3.2 Change from sire-MGS to sire-dam pedigree

The effect on sire variances and on breeding values was monitored by comparing results from scenario 2 and scenario 3.

Figure 1 shows the increase in sire variance in percent for protein yield, somatic cell count and direct calving ease for all countries participating in the international evaluation for production traits. Figure 2 shows a similar comparison Red Dairy Cattle.

In general, the figures show an increase in sire variance for almost all countries and traits. For protein yield, the average increase was 7.4%, which is somewhat higher compared to the findings of Van der Linde *et al.* (2005). The larger average change is to some extent because other countries were included. In addition, the completeness of pedigree differed somewhat between the present study and Van der Linde *et al.* (2005) and had an effect on the estimated sire variance.

Product moment correlations between and regressions of international breeding values across all bulls on each individual country scale were in general very high (Table 2).

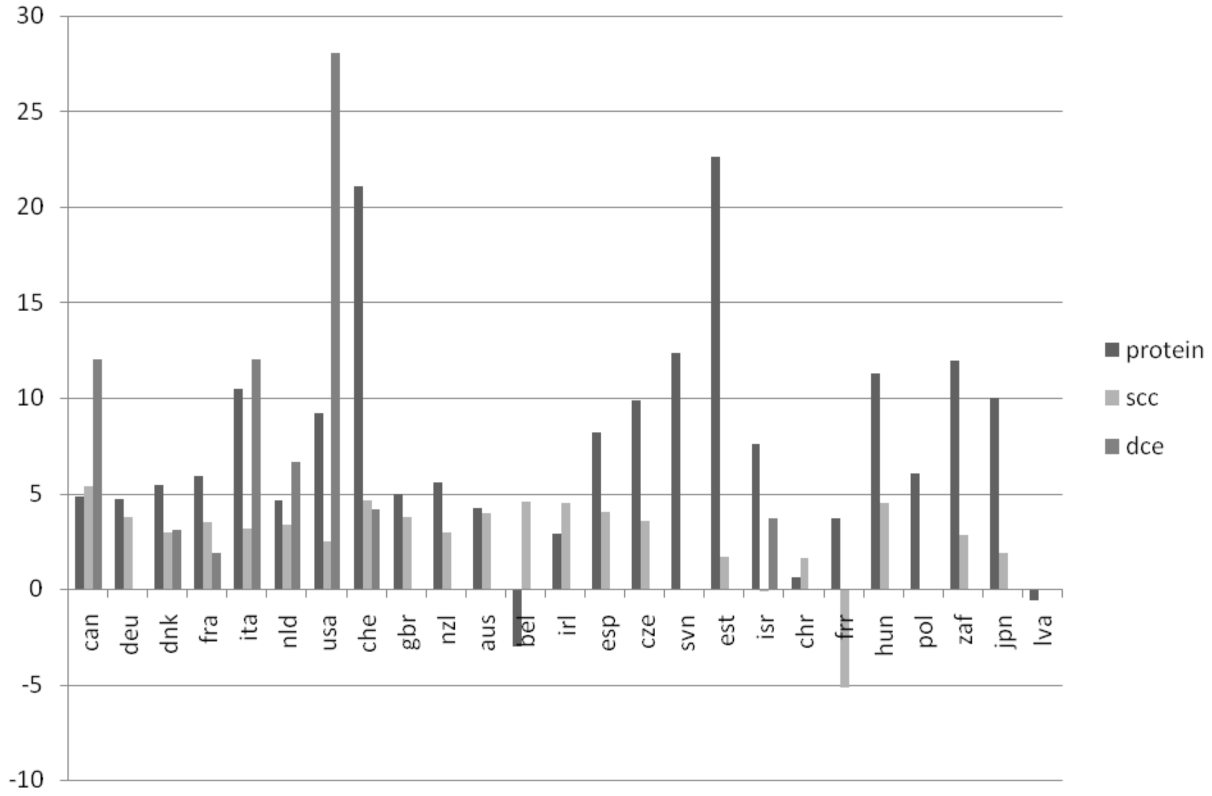

Figure 1. Increase (in %) of sire variance for Holstein for protein yield, somatic cell count (scc), and direct calving ease (dce) when changing from S-MGS to S-D pedigree in MACE (scenario 2 vs. 3).

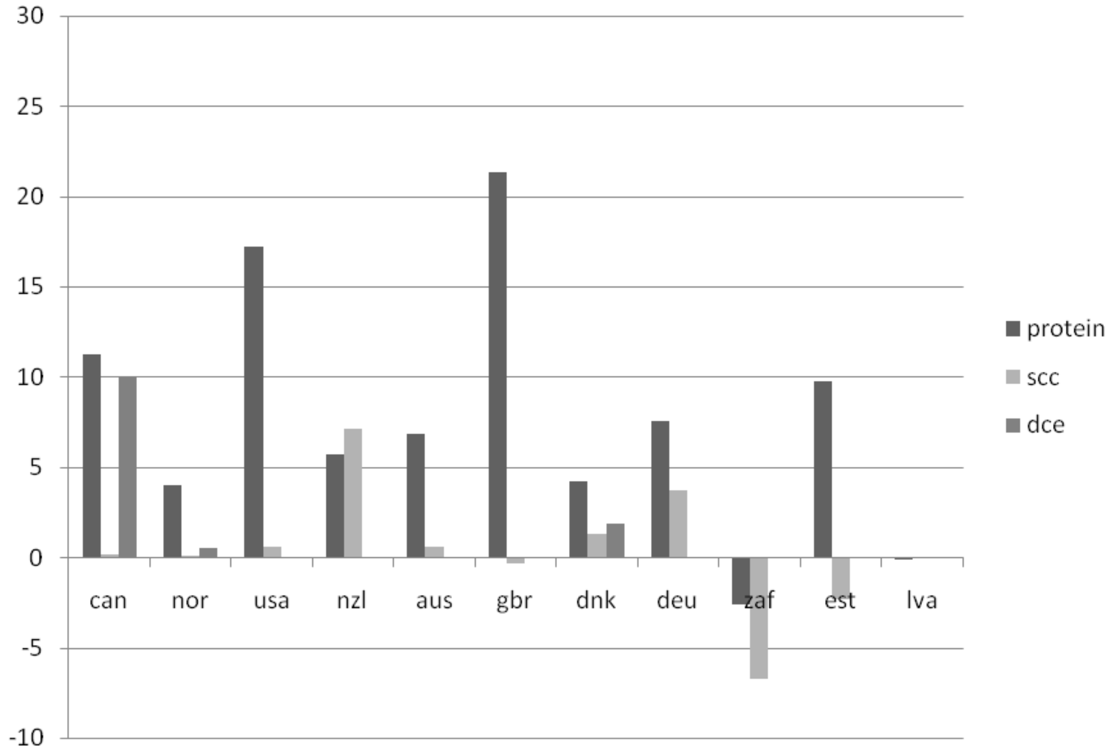

Figure 2. Increase (in %) of sire variance for Red Dairy Cattle protein yield, somatic cell count (scc), and direct calving ease (dce) when changing from S-MGS to S-D pedigree in MACE (scenario 2 vs. 3).

Table 2. Range of correlations and regressions across all bulls on each country scale (scenario 2 vs. 3).

	Holstein		Red Dairy Cattle	
	Correlation	Regression	Correlation	Regression
Protein	0.984-0.997	0.969-0.994	0.996-0.999	0.995-0.998
Somatic Cell	0.989-0.995	0.979-0.990	0.998-0.999	0.996-0.999
Direct Calving Ease	0.980-0.988	0.960-0.976	0.999-0.999	0.998-0.998

3.3 Depth of pedigree

Correlations between breeding values for protein yield obtained from scenario three and four across all bulls were above 0.999 on all country scales. For most countries the decrease in sire variance for Holstein, protein yield was less than one percent but for a few small populations the decrease was slightly larger (Figure 3).

Imposing the cut off year of 1970 for tracing of pedigree and forming phantom parent groups after that (scenario 4) is the same as currently used for correlation estimation and close to tracing of two generations of pedigree as is used for routine evaluation of the ten traits considered in the current study. However, tracing of pedigree as far back as possible is consistent with what is currently done for conformation and female fertility traits.

The CPU-time required for breeding value prediction for one trait and 25 countries in scenario three was four times longer than for scenario four.

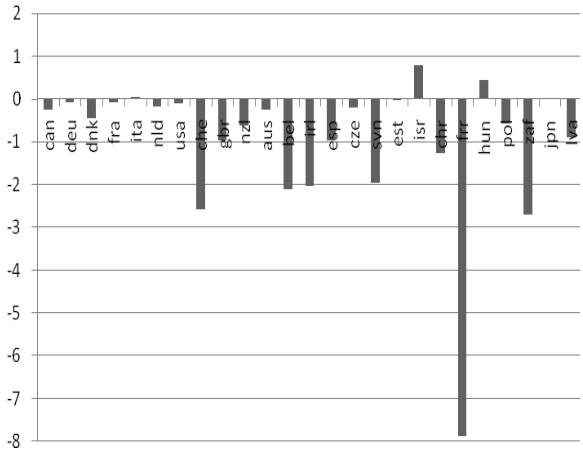

Figure 3. Change in sire variance for protein when reducing depth of pedigree (scenario 3 vs. 4).

3.4 Connectedness and change in reliability

Average numbers of common bulls were 172.0, 21.2, 63.5, 42.9, 56.1, and 18.8 for protein yield for Holstein, Red Dairy Cattle, Jersey, Simmental, Brown Swiss and Guernsey, respectively. These numbers show that there is a large difference in connectedness among the different populations participating in international genetic evaluation, which may result in differences between breed in the effect of changing pedigree structure on reliabilities. To picture the extremes in terms of connectedness the Red Dairy Cattle breed group was chosen for the analysis in addition to the Holstein.

More than 20,000 bull dams had two or more sons in the Holstein pedigree file (Table 3) and the corresponding number for RDC was 1,900 (Table 5). The difference in population structure becomes also clear from the intensity of use of bull sires: nearly 93% of the Holstein bulls had at least 9 half-sibs whereas this was only 78% for RDC (Tables 3 and 5).

Table 3. Number of Holstein sons born 1986 onwards per bull dam. Trait = Protein Yield.

Sons per bull dam	Number of dams	Percent of bulls
1	44,267	33.0
2	9,788	14.6
3	4,120	9.2
4	2,205	6.6
5	1,421	5.3
6-10	2,743	15.2
>10	1,282	19.1

Table 4. Number of Holstein sons born 1986 onwards per bull sire. Trait = Protein Yield.

Sons per bull sire	Number of sires	Percent of bulls
1	2,295	1.7
2	687	1.0
3	320	0.7
4	213	0.6
5	167	0.6
6-10	440	2.5
>10	1,019	92.9

Table 5. Number of Red Dairy Cattle sons born 1981 onwards per bull dam. Trait = Protein Yield.

Sons per bull dam	Number of dams	Percent of bulls
1	11,542	70.0
2	1,434	17.4
3	352	6.4
4	106	2.6
5	43	1.3
6-10	44	1.9
>10	6	0.5

Table 6. Number of Red Dairy Cattle sons born 1981 onwards per bull sire. Trait = Protein Yield.

Sons per bull sire	Number of sires	Percent of bulls
1	929	5.6
2	260	3.1
3	127	2.3
4	74	1.8
5	51	1.5
6-10	160	7.4
>10	376	78.3

The average change in reliability when changing from the sire-MGS model to the sire-dam model was across all bulls 1.16 for Holstein and 0.27 for Red Dairy Cattle (Scenario 2 vs. 3). However, when looking at sons of bull dams with several sons in the data the average change across countries was 2.20 for Holstein and 0.93 for Red Dairy Cattle. The corresponding values for bulls of sires with several sons in the data were 1.71 and 0.35 for the two breeds respectively. The figures show clearly a larger change in reliabilities for Holstein compared to the RDC breed which can be explained by the breed difference in connectedness (more common bull dams in Holstein).

The change in reliability by country is shown in Figures 4-7, and was larger for Australia, Ireland, New Zealand and Israel compared to the other countries. This can be explained by the fact that the average post-processed correlations for these countries were lower than for the other countries and the reliability for these countries did therefore have a larger opportunity to increase. Details about the post-processing procedure can be found on the Interbull home page (Interbull, 2009).

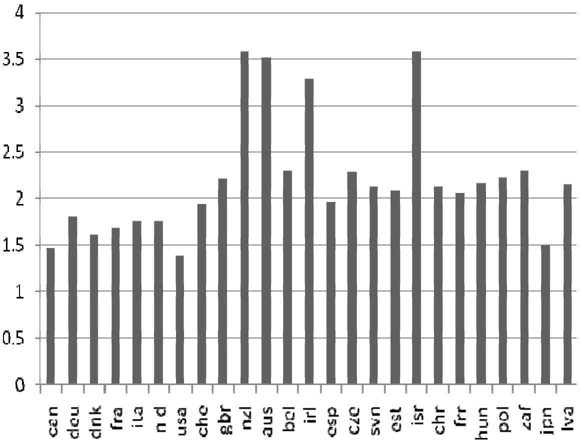

Figure 4. Average change in reliability for Holstein bulls whose mothers have several sons with proofs (scenario 2 vs. 3).

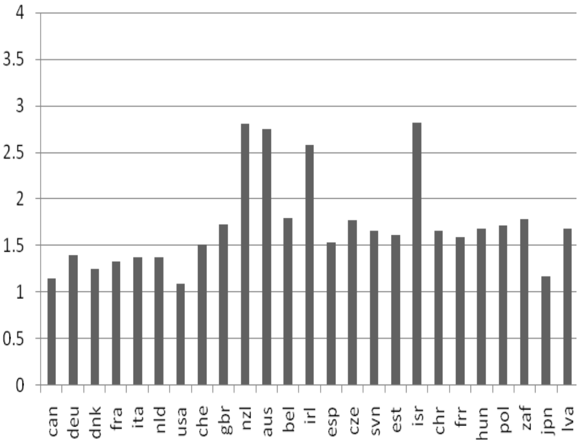

Figure 5. Average change in reliability for Holstein bulls whose sires have several sons with proofs (scenario 2 vs. 3).

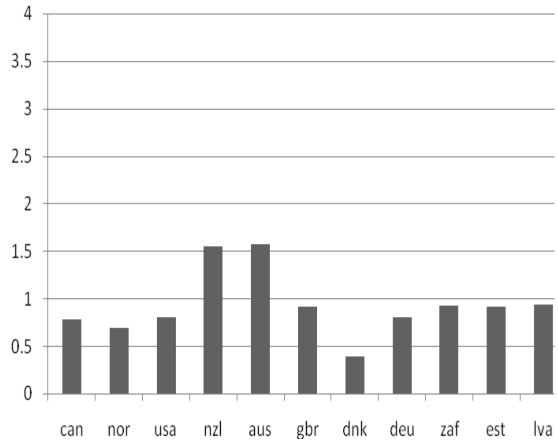

Figure 6. Average change in reliability for RDC bulls whose mothers have several sons with proofs (scenario 2 vs. 3).

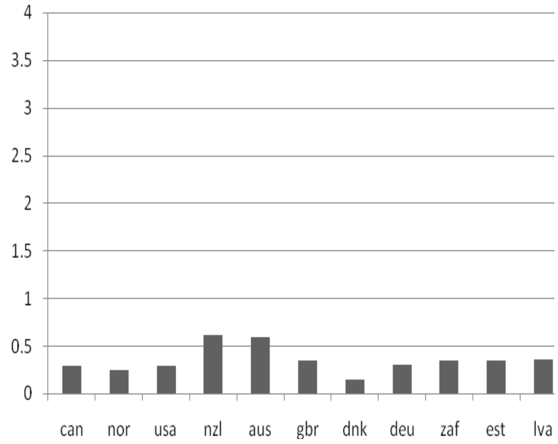

Figure 7. Average change in reliability for RDC bulls whose sires have several sons (scenario 2 vs. 3).

4. Conclusions

Implementation of sire-dam pedigree in Mace is feasible for different trait groups, for traits with high and with low heritabilities and for populations with different connectedness. The increase in reliability when changing from a sire-MGS pedigree to a sire-dam pedigree was largest for well connected populations.

Consistent sire-dam pedigree is a prerequisite for a successful implementation and it is necessary to solve the inconsistencies also on the dam side. A tool to verify pedigree (in line with the Interbull verification tool for breeding values) may be helpful.

The computing time decreased tremendously when pedigree was cut in 1970 and phantom parent groups were traced from then on. Product moment correlations between breeding values from analyses where pedigree was traced as far back as possible and breeding values from analyses where pedigree only was traced back to 1970 was very high.

References

De Jong, G. 2003. MACE – options for improvement. *Interbull Bulletin 30,* 112-116.

Fikse, W.F. 2008. Strategies for solving Mace MME. *Interbull Bulletin 38,* 146 – 151.

Harris B. & Johnson, D. 1998. Information source reliability method applied to Mace. *Interbull bulletin 17,* 31-36.

Interbull. 2009. http://www.interbull.org/documents/genetic_correlation_estimation_procedure.pdf. Accessed Feb. 17, 2009.

Klei, L. 1998. Solving MACE equations. *Interbull Bulletin 17,* 3-7.

Klei, L. & Weigel, K. 1998. A method to estimate correlations among traits in different countries using data on all bulls. *Interbull Bulletin 17,* 8-14.

Sigurdsson, A. & Banos, G. 1995. Dependent variables in international sire evaluations. *Acta Agric Scand., Section A, Anim. Sci. 45,* 209-217.

Sullivan, P. 1999. REML estimation of heterogeneous sire (co)variances for MACE. *Interbull Bulletin 22,* 146-148.

Van der Linde. C., de Roos, A.P.W., Harbers, A.G.F. & de Jong, G. 2005. Mace with sire-mgs and animal pedigree. *Interbull Bulletin 33,* 3–7.