Report from Workshop in Uppsala

J.B.M. Wilmink

NRS, P.O. Box 454, 6800 AL Arnhem, The Netherlands
E-mail: Wilmink.H@cr-delta.nl

F. Fikse

Interbull Centre, Dept. Animal Breeding and Genetics SLU, Box 7023, 75007 Uppsala, Sweden

Introduction

MACE contains many challenges! One of them is the estimation of the genetic (co)variance matrix. It was a great achievement when MACE was developed so that genetic correlations less than one between countries could be handled, but the estimation of the correlations has never been an easy task. The complexity has increased over time as more and more countries, and countries with weak genetic ties, join the evaluations at Interbull.

For that reason the Interbull Technical Committee (ITC) and a number of experts in the field of estimation of MACE genetic correlations had a workshop in Uppsala on 19 and 20 January 2004 (The expanded ITC workshop).

The objective of that meeting was two-fold:

1) Get a proper working procedure on how to handle rG-estimation given the current state of knowledge;
2) Identify critical issues in the current procedure that needs to be addressed, but that need further R&D.

The agenda for the expanded ITC workshop

Three topics were discussed on the first day. These were: I) data edits, pedigree and genetic groups, II) estimation procedures and III) post processing of estimates.

Every topic was introduced by a paper and discussion points that was presented by our colleagues of the Interbull Centre. The points were discussed in three groups, results were presented and, if possible, some conclusions were drawn.

The second day was for introductions on special topics looking for new and fresh looks on the problem of genetic correlation estimation and for meetings of the ITC and the Interbull Scientific Advisory Committee (SAC). The following issues were presented:

- ✓ Can we make estimation of rG robust and simple?
- ✓ Direct estimation of a reduced rank genetic correlation matrix for international evaluations;
- ✓ Application of structural equation modeling in international genetic evaluations.
- ✓ Pseudo-REML procedure to estimate all covariances simultaneously for multiple-trait MACE;
- ✓ Estimation of genetic correlations when evaluations of both parents are available;
- ✓ h2 from data or countries?
- ✓ SAC's critical issues with respect to correlation estimation.

The participants of the expanded ITC workshop were:

Georgios Banos, Freddy Fikse, Zengting Liu, Raphael Mrode, Pete Sullivan, Paul van Raden, Vincent Ducrocq, Hossein Jorjani, Gerrit Kistemaker, Jette Jakobsen, Gerben de Jong, Per Madsen, Mike Goddard, Thomas Mark, Esa Mäntysaari, Larry Schaeffer, Stephanie Minery, Tom Lawlor, Jan Philipsson, Ulf Emanuelson, Hans Wilmink (chair).

Results

Data Edits

Topic I was on data edit, pedigree and genetic grouping related to estimation of genetic correlations. Discussed questions were:

- ✓ On data edit: should all bulls or a sub set of bulls be used when possible? And how should sub sets be selected?
- ✓ How many generations of pedigree?
- ✓ Grouping: should phantom grouping be random or fixed?

Results were:

- ✓ Subsetting is needed, but as few as possible estimates for the same correlation;
- ✓ It is better to have more bulls than more countries: try to represent the population to avoid selection bias;
- ✓ Pedigree: take as many generations as possible, but be consistent with the EBV estimation;
- ✓ It is better to fix a time period, than the number of generations; This may be different per trait;
- ✓ Genetic groups should be considered to account for genetic trend. If the animal model is used, there will be no difference between fixed and random. Phantom groups should not be too small.

Algorithms and program packages

Discussion points were:

- ✓ Should groups in genetic correlation estimation be fixed or random?
- ✓ What package to use? EM-REML or AI-REML;
- ✓ Are Structural Models (SM) ready for routine application?

Results were:

- ✓ Groups should be considered as random; there is no need to change the current approach;

- ✓ AI-REML is preferred. This will give SE, it is faster and have more precise maximum and can be modified to assume constant heritability;
- ✓ Structural Models: Advantages are that fewer parameters need to be estimated; no need to analyse all countries together or to re-estimate parameters, easy to compute and poorly connected countries can be considered. Questions are: how to define the axes, what is the behaviour when correlations are low, heritabilities must be fixed. Structural Models should be complementary to classical estimation methods;
- ✓ There was a discussion on Structural Models related to principal components/ factor analysis. Structural Models is the presentation of genetic correlations by distances, which are by definition positive. SM may not represent all genetic correlation structures well. It would be better to use principal component or factor analysis to represent the N*N correlation matrix in a smaller space.

Post processing

Discussion points on the post processing were:

- ✓ Use only pure estimates or also other information sources;
- ✓ Which information sources should be used then?
- ✓ How to combine estimates and expectations?

Results were:

- ✓ Use all information which is available
- ✓ Information sources are: the pure estimates, other breeds, own expectations; previous estimates may be used to achieve consistent results but is principally not needed;
- ✓ A clear documented method is needed to combine the estimates; a sort of Bayesian approach. Post processing to get the proper correlation estimates is needed.

The results were discussed in the ITC. A summary of the conclusions is in Table 1.

Table 1. A summary of the conclusions by the ITC of the expanded ITC workshop in Uppsala, 2004.

	Production	Conformation	Udder Health	Longevity	Calving traits	ITC recommendations based on workshop discussions
Estimation procedure: Sigurdsson et al. 1996 Klei & Weigel, 1998	SBP96	KW98	SBP96 KW98	KW98	KW98	Production should move to KW98.
Estimation frequency	First time a population joins evaluations. Thereafter, whenever there is a change of N-GES.	First time a population joins evaluations. Thereafter, at each test run.	First time a population joins evaluations. Thereafter, at each test run (*).			Keep as it is. At least until production uses KW98, when a new discussion should be taken.
Input data	De-regressed national EBV for sub-set(s) of AI bulls.	De-regressed national EBV for sub-set(s) of AI bulls that have pedigree information in production data file.	De-regressed national EBV for sub-set(s) of AI bulls that have pedigree information in production data file.	De-regressed national EBV for sub-set(s) of AI bulls that have pedigree information in production data file.	De-regressed national EBV for sub-set(s) of AI bulls that have pedigree information in production data file.	Not discussed during the workshop.
Data edit: # of daughters			50 daughters (CM)		50 daughters	Not discussed during the workshop.
Country sub-setting: # of countries	(*) By SBP96: All 2-country combinations; All 3-country combinations; All 4-country combinations; Some 5-country combinations; Some 6-country combinations.	(*) By KW98: HOL: Some 7-country combinations (two 3-country combinations + USA); OTH: No country sub-setting.	(*) By SBP96: All 2-country combinations; Some 3-country combinations (HOL: countries with a change in N-GES). (*) By KW98: HOL: Some 7-country combinations (two 3-country combinations + USA); OTH: No country sub-setting.	(*) By KW98: HOL: 14 countries have participated for pilot. Correlations estimated for all countries simultaneously.	(*) By KW98: HOL: 10 countries have participated for pilot. Correlations estimated for all countries simultaneously.	Use as many countries as possible. Estimation for Holstein production should use a similar procedure as Holstein conformation.
Bull sub-setting: Use of only common bulls	HOL: For large populations and or large country sub-sets					
Bull sub-setting: Use of common bulls and their ¾ sibs	OTH: For small populations and / or small country sub-sets	Yes	Yes	Yes	Yes	Common bulls **and** their ¾ sibs should be used.

Pedigree: # of generations	(*) SBP96: 2 generations	(*) KW98: As far back as possible	(*) SBP96: 2 generations (*) KW98: As far back as possible	(*) KW98: As far back as possible	(*) KW98: As far back as possible	The pedigree should be traced as far back as possible, but parents born before 1970 should be set to missing.
Tracing of pedigree	(*) SBP96: Trait wise within trait group	(*) KW98: Across traits within trait group	(*) KW98: Across traits within trait group	(*) KW98: Trait wise within trait group	(*) KW98: Trait wise within trait group	The pedigree must be traced separately for each trait.
Phantom parents: # of groups	(*) SBP96: 3 groups (S, MGS and MGD)	(*) KW98: Variable: Based on national origin, breed, birth year and selection path (S, MGS and MGD).	(*) SBP96: 3 groups (S, MGS and MGD); (*) KW98: Variable: Based on national origin, breed, birth year and selection path (S, MGS and MGD).	(*) KW98: Variable: Based on national origin, breed, birth year and selection path path (S, MGS and MGD).	(*) KW98: Variable: Based on national origin, breed, birth year and selection path path (S, MGS and MGD).	The approach in KW98 should be used for all traits.
Phantom parents: Minimum group size	10	15	30	30	30	Keep as it is, but increase to 15 when production has moved to KW98.
Starting values	Previously published correlations	Previously published correlations	0.85	0.85	0.70	Previously published correlations.
Convergence	(*) By SBP96: After 150 iterations	ALL: $\lambda < 10^{-4}$ or 10000 iterations; $$\lambda_{ij} = G_{ij} \big/ \sqrt{R_i R_j}$$	(*) By SBP96: After 400 iterations. (*) By KW98: HOL: $\lambda < 10^{-4}$ or 5000 iterations; AYR: $\lambda < 10^{-6}$ or 8000 iterationss OTH: $\lambda < 10^{-7}$ or 10000 iterations. $$\lambda_{ij} = G_{ij} \big/ \sqrt{R_i R_j}$$	(*) By KW98: HOL: $\lambda < 10^{-6}$ or 10000 iterations $$\lambda_{ij} = G_{ij} \big/ \sqrt{R_i R_j}$$	(*) By KW98: HOL: $\lambda < 10^{-5}$ or 10000 iterations $$\lambda_{ij} = G_{ij} \big/ \sqrt{R_i R_j}$$	Should be equal for all traits and no change in individual correlations between iterations should be larger than 10^{-4}.
Post processing: Information sources In addition to pure current estimates	Previous estimates, own expectation	AYS: Estimates from HOL	HOL: Previous estimates, Own expectation OTH: As above + HOL			No decision made.
Post processing: Bending	Yes (when necessary)	Yes (when necessary)	Yes (when necessary)	Yes (when necessary)	Yes (when necessary)	Keep as it is.