Interbull Scientific Advisory Committee (SAC)

Annual report (2004-2005) to the Interbull Steering Committee (SC)

Interbull SAC: *Vincent Ducrocq, Mike Goddard, Larry Schaeffer, Georgios Banos (convener)*

Another year of increased Interbull activity has elapsed and SAC congratulates the Interbull Centre for all their hard work and thanks them for good communication and cooperation.

The objective of this report is to inform the Interbull SC of our activities during 2004-2005 and provide views on pertinent scientific issues.

Jan Philipsson, Freddy Fikse and Georgios Banos met during the 2004 Interbull meeting in Sousse, Tunisia. Jan and Freddy suggested that **data quality**, in particular **data validation** methods, was the key issue on which the views of the SAC members were sought.

BACKGROUND

At present, Interbull data validation is primarily based on:

1. The three trend validation methods.
2. A proposed method to check the stability of the genetic variance estimate ("Proposal for procedure to validate Mendelian sampling variances").

These two methods along with other possible approaches are discussed next.

TREND VALIDATION METHODS

Briefly, Method 1 compares trends from first-lactation and all-lactations evaluation models. Method 2 examines the changes in yearly daughter deviations over time. Method 3 examines the weighted regression of the most recent proof on proportion of new daughters and proof four years ago. In all case, deviations of magnitude larger than 2% (1% if proofs are predicted transmitting abilities) of the genetic standard deviation indicate that trends are biased.

These methods were developed for milk production traits and (mostly) lactation animal models, and have since been implemented to other traits and model results. Their relevance is discussed next.

1) Suitability of current methods for test-day models for the evaluation of milk production traits.

Method 1 should be applicable, with caution when lactations are treated as different traits.

Method 2 depends on daughter deviations, whose definition may be ambiguous for test-day models. In order to alleviate this ambiguity and produce a uniform definition, one should be looking for an un-regressed measure of daughter performance adjusted for all fixed effects. For example, Vincent Ducrocq's paper at the 2004 Interbull meeting dealt with just that.

Briefly, the weight (W) and daughter deviation (D) can be defined so that the expression $[(\mathrm{diag}(W) + \sigma^2_e/\sigma^2_g A^{-1}) = \mathrm{diag}(W)^* D]$ yields the same sire solution as the national genetic evaluation model. Test-day models may then generate daughter deviations per cow and day of lactation that could be combined into a single measure per sire daughter and then per sire.

Method 3 should be applicable, using the published (official) proofs.

2) Suitability of current methods for <u>new traits</u> (conformation, udder health, longevity, calving traits).

Method 1 is probably not applicable because a) there is no equivalent to first lactation for longevity, b) conformation data are usually from one lactation only, and c) first and later parities are often considered as different traits of fertility, calving performance, mastitis resistance etc.

Regarding method 3, it may be applicable to some of these traits but not to longevity. In the latter, when records from second crop daughters become available, the trait definition of first crop daughters has changed, since they may have now had longer productive life. So the contribution of the second crop to the latest proof may be masked by the updated contribution of the first.

Furthermore, in some cases (mastitis, longevity, calving traits), data and breeding values aren't normally distributed. This would affect methods 2 and 3. The current confidence intervals for tolerance levels are constructed based on the implicit assumption that residuals are normally distributed. Empirical confidence intervals for method 3, obtained with non-parametric approaches (e.g. as per Weller et al, 2003 Interbull meeting) might be more suitable in this regard.

When threshold models are used for the national genetic evaluation of some of these traits (e.g. calving performance) the current trend validation methods are non-applicable. In particular, weights on daughter deviations do not fully reflect the true daughter contribution. The basic principle of deriving weights and daughter deviations, as describe above for method 2 for milk traits, should also apply here. When possible, exact reliabilities could be calculated by direct inversion of the coefficient matrix and used to compute daughter contributions after accounting for pedigree.

3) Appropriate tolerance levels.

Current levels (2% or 1%) were somewhat arbitrarily set for milk production traits. Different tolerance levels are probably needed for different traits, dependent on the variation of these traits. These might be also arbitrary, to some extent, but they would pertain to the trait in question.

Another caveat with some functional traits is that, because of their genetic correlation with milk yield, trends derived from uni-variate models for genetic evaluations are likely to be affected by selection bias. This wouldn't be the case if multi-variate models, including milk yield, are used for the evaluation of functional traits.

Finally, if the regression is non-significantly different from zero (parametric or non-parametric testing) tolerance levels should be irrelevant.

PROPOSED VALIDATION OF MENDELIAN SAMPLING VARIANCE

In general, there appears to be merit in this approach. Bootstrap is certainly suitable when dealing with small population sizes. The optimum number of samples still needs to be determined, though.

The proposed method is based on national data and, apparently, reliability estimates calculated in each country separately. Reliability estimation procedures have probably converged to some extent, especially for certain model families; however, there are still differences in the evaluation systems of various countries.

Mendelian sampling (m) may also be used in several other ways, such as looking at average m and m'm within birth year, age at calving, season of calving, herd size, trait level etc, and testing for differences between groups. Evidence of significance for any of these strata may suggest that the model does not remove all bias. The same exercise can be repeated with the mixed model residuals ($e=y-Xb-Zu$) instead of m. Furthermore, e and m can be also computed from reduced data (e.g. excluding records from the last year) and compared with values using all data available. This will indicate how fixed effect solutions may affect results when lactations are still in progress.

OTHER APPROACHES

1. Looking at data subsets

The key to assessing the validity of any genetic evaluation would be to ensure that estimated breeding values (EBV) of bulls (especially of those whose semen will be broadly marketed) accurately predict future daughter performance. A useful test would be to re-compute today's genetic evaluations after having excluded data from the last four years and compare them with adjusted daughter records (or daughter deviations) in the last four years or with currently official EBV. Regression of daughter deviations on reduced data EBV should be near unity. Differences between reduced data and official EBV should average zero, be independent of the former and have variance proportional to the change in reliability. Conceptually, this is similar to method 3 without having to wait for four years to get the results; also there are no potential problems from changing evaluation models between first and second crop. The method can be applied at the Interbull Centre and the procedure may be repeated for any time-window in the national data-base.

2. Robust models

Another approach is to try to develop international genetic evaluation models robust to incorrect trends. This would involve models that include a country-by-birth year effect. Dependent variables and weights would then be daughter deviations and effective daughter contributions, respectively, calculated within year in each country (similarly to trend validation method 2). Simulation studies conducted in the framework of the European "PROTEJE" program suggested that such models may recover biases emanating from erroneous trends. This idea is worth investigating further to ensure results are indeed robust to incorrect trends, regardless of what causes the bias.

3. Data-mining

Work on data-mining applications to genetic evaluation data control is currently in progress. The third report will be presented at the 2005 Open meeting. A platform is being developed to automatically apply several data-mining tests to all countries and traits in the international evaluation. There is still some way to go before one can fully understand and interpret all data-mining associations, but certain clear results have been derived that may provide useful information for the assessment of data quality.

FINAL REMARKS

The need to develop methods and systems that can guarantee the unbiasedness of international genetic evaluations has been clearly articulated on many occasions. The methods reviewed and outlined in this report may fall into any of the following three categories:

1. Assessing data quality with the view to discard data that do not meet certain standards (*trend validation, mendelian sampling, data-mining*).
2. Assessing the suitability of genetic evaluation models for data analysis (*data subsets*).
3. Developing models that are robust to input data (*robust models*).

Each method should be evaluated and tested in its own merit. There may be possibilities to combine two or more complementary methods. The clear objective should be to improve the quality of the international genetic evaluations and enhance the acceptability of results in the global animal genetics sector.

May 23, 2005

On behalf of the Interbull SAC

Georgios Banos
Convener