Prior Genetic Correlations and Non-Measured Traits

T. Mark[1], P.G. Sullivan[2], W.F. Fikse[3] and P.M. VanRaden[4]

[1]The Royal Veterinary and Agricultural University, Frederiksberg, Denmark. [2]CDN, Guelph, Canada. [3]Interbull Centre, Uppsala, Sweden. [4]AIPL, Beltsville, United States

Introduction

Current international genetic evaluations are based on across-country genetic correlations (r_G). These r_G are estimated r_G that have been "post-processed" and may be influenced by prior expectations. Current post-processing rules are largely based on expert intuition. Applying similar structural models as Rekaya *et al.* (2001) to predict prior r_G seems more desirable as it allows simultaneous consideration of several explanatory effects and because it is less subjective.

The relative weight of the prior increases as the precision of the estimated r_G decreases. In some cases, estimated r_G have very low precision (e.g., Mark *et al.,* 2005a), and in such cases, r_G may be based almost entirely on prior information. Mark (2005) showed that international genetic evaluations can be obtained for "non-measured" traits, based on correlated information from measured traits included in current Interbull evaluations, if suitable r_G are available among the measured and non-measured traits. Examples of non-measured traits are milk production in Brazil and clinical mastitis in the United States.

The aim of this paper is to develop multiple regression models to predict prior r_G and to illustrate how these priors can be used in MACE inferences for both measured and non-measured traits. Emphasis will be on, but not limited to, milk yield for Holsteins.

Material

National genetic evaluation results for Holstein milk yield used in the March 2006 Interbull test evaluation were considered in this study. Variables potentially explaining variation associated with estimated r_G were obtained from three sources, and accordingly, they could be grouped into 1) climatic variables, 2) production system indicators and 3) national genetic evaluation descriptors.

The climatic variables were available from the Danish Meteorological Institute and were measured as the average monthly value during 1931 to 1960 in the capital city (Cappelen and Jensen, 2001). For Germany-Austria, Slovenia and Switzerland, the average wind speed was set to the overall average of other countries due to missing information for these countries. The variables considered here were country averages of temperature (Celsius), range in temperature (from coldest to warmest month), rainfall (mm), range in rainfall, humidity (pct), range in humidity and wind speed (Beaufort).

Production system indicators were available from ICAR's yearly enquiries. The most recent statistics were taken from each country. A weighted average was used for Denmark, Sweden and Finland (weighted according to bulls included in the Interbull test evaluation). Holstein data were used when available; otherwise, statistics for all cows were used. The indicators that were considered here were: average milk yield (kg) fat and protein contents (%) from national milk recordings. Furthermore, Australia, Ireland and New Zealand were treated as countries with grassing during the whole year like in the current post-processing of r_G.

National genetic evaluation descriptors were taken from the forms that the national genetic evaluation units use to describe their evaluation models (Form GE; available on Interbull's homepage). The descriptors that were considered here were: heritability, parities included (>3 treated as 3 parities), test-day records or not, repeatability model or not, analysed simultaneously with biologically different traits or not and whether a fertility measure was included as an explanatory variable or not. Furthermore, the number of bulls with an evaluation in each of two countries (common bulls) was available.

Methods

Across-country genetic correlations were estimated using EM-REML (Klei and Weigel, 1998) as in official Interbull evaluations. Multiple linear regression was used to predict prior r_G for milk yield from explanatory variables. Power functions of estimated r_G (r^k_G, k=1, 2, 3, 4, 5, 6, 7) were considered as the dependent variable, in attempts to reduce skewness in the distribution of observations. The explanatory variables were expressed as ratios or binary variables. For continuous variables, a ratio was calculated so that the highest of the two country averages was in the denominator. Hence, $0 < \text{ratio} \leq 1$, and a high ratio always indicated that the variable in question was similar in the two countries. Likewise, a binary class variable was set equal to 1 if both countries belonged to the same class (e.g., both countries considered the same number of parities); otherwise, it was set equal to 0. The number of common bulls was used as is. All variables were constructed so that the regression coefficient was expected to be positive. The best model for r^k_G was selected based on Mallow's C(p) with the restriction that the regression coefficient associated with a variable must not be negative if the variable were to be included in the model. This restriction was imposed to make sure that the derived prediction formula would be biologically meaningful.

Results and Discussion

Prediction of prior genetic correlations. The dependent variable that gave the best fit was r_G raised to the power of 5 (6 was equally good). The percentage of variation explained by the best model for r^5_G was 47.1% whereas it was 44.6% for r_G. The best model for r^5_G, based on Mallow's C(p), was:

$$r^5_G = \mu + b_1(\text{milk}) + b_2(\text{grass}) + b_3(\text{wind}) + b_4(\text{temp}) + b_5(h^2) + b_6(\text{par}) + b_7(\text{CB}) + \varepsilon \quad [1],$$

where milk, grass, wind, temp, h^2 and par were the ratio for milk yield, grassing, wind speed, temperature, heritability and number of parities, respectively, for the two countries corresponding to the estimated correlation, CB is the actual number of common bulls, and ε is the residual. The CB was used both as weighting factor in the analysis and as an explanatory effect because it was expected to be related to both the precision of the estimated r_G and may explain the actual level of r_G. The latter could be the case if it indicated similarity of the production systems or because low CB caused underestimation of r_G (Sigurdsson *et al.,* 2006). Using CB as a weighting factor gave a better model fit than using either the square root of CB or equal weights regardless of whether CB was included as explanatory effect or not. Akaike Information Criteria and backward selection yielded the same best model as Mallow's C(p). The estimated regression coefficients for Model [1] are presented in Table 1.

Model [1] tended to shrink prior r_G towards the average r_G. Thus, the standard deviation of estimated r_G was 0.11, whereas the standard deviation of prior r_G was 0.056. For countries with relatively low average r_G, such as the South African Republic and New Zealand, the average prior r_G was higher than the average estimated r_G (Table 2). This may indicate that the model did not describe real differences in production systems or that the model partly corrected for biased r_G. Certainly, the model may be improved. For instance, it may be possible to obtain statistics for the average herd size or for the average months on grass as well as variables highlighted by Zwald *et al.* (2003), such as peak milk yield. Member countries may provide this information. Also, in retrospect, we realised that we should have allowed a squared temperature term to be included in the final model although the regression coefficient was negative (p=0.03) because performance in very cold and very warm environments could be similar.

The first four effects in Model [1] (i.e. milk, grass, wind and temp) can be interpreted as causing true genotype by environment interaction (G×E) whereas the remaining effects are not clearly associated with G×E. Heritability differences can be due to data quality and national evaluation models. The number of parities included in evaluations is often a pragmatic choice that is not necessarily related to the production system. As discussed above, CB is more difficult to interpret, but it is probably not fully associated with true G×E.

It may be desirable to eliminate sources of variation not fully related to G×E when generating prior r_G, at least for non-measured traits. In Table 3, prior r_G for NZL is presented both using Model [1] with the parameter estimates from Table 1 and using Model [2]: $r^5_G = \mu' + b_1(\text{milk}) + b_2(\text{grass}) + b_3(\text{wind}) + b_4(\text{temp}) + \varepsilon$, where $\mu' = \mu + 0.144(1) + 0.042(1) + 0.225(990)$ and variance = var(μ); i.e., h^2, par and CB have been fixed at their maximum values (1, 1 and 990, respectively). Forcing variables not fully associated with G×E to be equal in two countries resulted in noticeably higher prior r_G for New Zealand (Table 3). Weigel *et al.* (2001) used a uniform method on raw data to remove statistical artefacts, which may be better in theory but is challenging in practice. The average r_G for New Zealand, with countries included in this study and also in the study of Weigel *et al.* (2001), were 0.67, 0.86 and 0.87 for priors using Model [1], priors using Model [2] and estimates from Weigel *et al.,* respectively. This verifies that Model [2] worked as intended for New Zealand.

Use of prior genetic correlations. One feature of the outlined procedure to obtain prior r_G is that standard errors are available. These can be used to give appropriate weights to each specific prior relative to the estimated r_G. This and the fact that each prior mean is obtained in a more comprehensive way than for current Interbull practice should help to improve post-processing of estimated r_G, which is based on a simple weighted average of prior and estimated r_G (Mark *et al.,* 2005b). Finally, this approach gives a framework that enables prediction of prior r_G for non-measured traits.

Mark (2005) showed how knowledge of prior r_G for non-measured traits enables predictions of international breeding values for such traits. Mark *et al.* (2006) showed that more efficient selection for clinical mastitis in countries without direct mastitis records can be achieved by using this method. However, in that study, too few r_G were available for clinical mastitis to use the multiple regression technique to predict prior r_G. Instead, the r_G for milk somatic cell in the same two countries was used to guide the r_G for clinical mastitis. Different approaches to predict prior r_G (or at least different explanatory variables) are probably needed for different trait groups.

Conclusion

Prior across-country r_G for milk yield were generated with reasonable precision (standard error ≈ 0.3) using multiple regression. Such priors can be used to improve post-processing of estimated r_G among measured traits and to predict international breeding values for non-measured traits. Prior r_G increased noticeably (0.2 on average for New Zealand) when sources of variation not associated with true G×E were eliminated. Further investigation is needed to determine which priors should be preferred in international evaluations.

References

Cappelen, J. & Jensen, J.J. 2001. *Teknisk rapport 01-17. DMI.*

Klei, L. & Weigel, K.A. 1998. *Interbull Bulletin 17,* 8.

Mark, T. 2005. *Doctoral thesis no. 93. SLU.*

Mark, T., Fikse, W.F., Sullivan, P.G. & VanRaden, P.M. 2006. *Proc. 8th WCGALP.*

Mark, T., Madsen, P., Jensen, J. & Fikse, W.F. 2005a. *J. Dairy Sci. 88,* 3303.

Mark, T., Madsen, P., Jensen, J. & Fikse, W.F. 2005b. *J. Dairy Sci. 88,* 3290.

Rekaya, R., Weigel, K.A. & Gianola, D. 2001. *J. Dairy Sci. 84,* 1525.

Sigurdsson, A., Banos, G. & Philipsson, J. 1996. *Acta Agric. Scand. (A) 46,* 129.

Weigel, K.A, Rekaya, R., Zwald, N.R. & Fikse, W.F. 2001. *J. Dairy Sci. 84,* 2789.

Zwald, N.R., Weigel, K.A., Fikse, W.F. & Rekaya, R. 2003. *J. Dairy Sci. 86,* 1009.

Table 1. Parameter estimates (b), their standard errors (SE) and significance level (p) for Model [1].

Variable	b	SE	p(b=0)
μ	–0.586	0.11	0.000
milk	0.491	0.11	0.000
grass	0.103	0.03	0.001
wind	0.235	0.07	0.001
temp	0.187	0.07	0.013
h^2	0.144	0.07	0.032
par	0.042	0.02	0.042
CB ($\times 10^{-3}$)	0.225	0.04	0.000

Table 2. Mean bias and mean squared error (MSE) of predicted priors on transformed and original scale for each country[1].

	Transformed		Original scale	
Country	Bias	MSE	Bias	MSE
AUS	–0.041	0.017	–0.019	0.006
BEL	0.056	0.023	0.035	0.007
CAN	–0.046	0.025	–0.005	0.005
CHE	–0.032	0.036	–0.001	0.007
CZE	0.188	0.051	0.099	0.013
DEU	0.053	0.016	0.034	0.006
DFS	–0.037	0.017	–0.009	0.003
ESP	–0.062	0.025	–0.014	0.004
EST	0.108	0.030	0.102	0.028
FRA	–0.094	0.036	–0.025	0.006
GBR	0.029	0.015	0.021	0.004
HUN	0.044	0.017	0.027	0.005
IRL	–0.058	0.030	–0.013	0.008
ISR	0.094	0.021	0.065	0.010
ITA	0.007	0.016	0.016	0.005
JPN	–0.103	0.038	–0.028	0.007
NLD	–0.048	0.021	–0.009	0.004
NZL	0.106	0.016	0.087	0.012
POL	–0.057	0.027	–0.014	0.005
SVN	0.016	0.033	0.028	0.012
USA	–0.015	0.021	0.008	0.005
ZAF	0.178	0.037	0.148	0.027

1) Mean bias original scale: $\Sigma(r_G\text{prior} - r_G)/n$; transformed scale: $\Sigma(r^5_G\text{prior} - r^5_G)/n$

Table 3. Estimated genetic correlations (r_G) with New Zealand, prior r_G obtained without (prior) and with (prior+) forced harmonisation of arbitrary differences between countries; prior mean ± SD of 100,000 samples from Model [1] and [2], respectively.

Country	r_G	Prior	Prior+
AUS	0.85	0.81 ± 0.14	0.88 ± 0.06
BEL	0.66	0.64 ± 0.36	0.86 ± 0.08
CAN	0.66	0.71 ± 0.28	0.87 ± 0.08
CHE	0.72	0.61 ± 0.39	0.85 ± 0.08
CZE	0.61	0.52 ± 0.46	0.83 ± 0.09
DEU	0.58	0.68 ± 0.32	0.86 ± 0.07
DFS	0.70	0.62 ± 0.38	0.85 ± 0.08
ESP	0.68	0.57 ± 0.43	0.84 ± 0.08
EST	0.47	0.69 ± 0.31	0.87 ± 0.07
FRA	0.74	0.67 ± 0.33	0.86 ± 0.07
GBR	0.71	0.74 ± 0.24	0.87 ± 0.07
HUN	0.63	0.61 ± 0.39	0.85 ± 0.08
IRL	0.81	0.87 ± 0.10	0.93 ± 0.05
ISR	0.54	0.55 ± 0.44	0.83 ± 0.09
ITA	0.67	0.65 ± 0.35	0.85 ± 0.08
JPN	0.70	0.55 ± 0.44	0.84 ± 0.08
NLD	0.69	0.67 ± 0.32	0.85 ± 0.08
POL	0.69	0.70 ± 0.30	0.89 ± 0.06
SVN	0.75	0.73 ± 0.27	0.88 ± 0.06
USA	0.66	0.76 ± 0.21	0.84 ± 0.09
ZAF	0.63	0.46 ± 0.50	0.84 ± 0.09