Accounting for Foreign Information in Genetic Evaluation

BERNARD BONAITI and DIDIER BOICHARD
Institut National de la Recherche Agronomique,
Station de Génétique Quantitative et Appliquée,
F 78352 Jouy-en-Josas Cedex, France

ABSTRACT

This note presents a method to include foreign information in a national genetic evaluation system based on a standard single trait repeatability model. When available, daughter yield deviation (DYD) is easily incorporated by generating virtual daughters with one performance equal to the converted DYD. The other parents of these virtual daughters are assumed to be unknown and put into a special group, which estimate is simply related to the a-factor of conversion formula. When only EBVs are available, the performance to be attrituted to virtual daughters is a "deregressed" EBV, which expression is given. The method is illustrated by the inclusion of the US Holstein and Brown Swiss information in the French evaluation system.

INTRODUCTION

In international breeds, a large proportion of genes may originate from foreign countries, when semen, embryos, or live animals are imported. As these foreign animals are usually strongly selected, omitting the foreign information would bias the national genetic evaluation. Foreign information includes pedigree, performances, and estimated breeding values (EBV). The pedigree information may be easily accounted for through the relationship matrix. To account for different genetic levels between base populations in different countries, groups of unknown parents may be defined according to the country of origin, in addition to the usual criteria, as sex, birth year, or strain. However, because the imported animals are not average animals randomly sampled in the foreign population, it is also necessary to account for their individual genetic merit. Raw performances cannot be used, unless a joint evaluation is performed (1) with a complete data set of both countries. In most situations, it is desirable to include only informative foreign animals in the national evaluation. These selected foreign ancestors are characterized by a foreign EBV and by related information (reliability, daughter yield deviation (DYD)...). This note presents a method to include foreign information in a national genetic evaluation system.

METHODS

Let us consider a standard single trait repeatability model, including fixed effects, a vector of additive genetic effects with variance $\mathbf{A}\, h^2 \sigma^2$, a vector of permanent environmental effects with variance $\mathbf{I}\, (\rho-h^2)\, \sigma^2$, and genetic groups defined for unknown parents only (6).

Incorporating DYD information

The DYD is a direct measure of the daughters superiority of a sire and does not include pedigree information, in contrast to EBVs. First it should be converted into the importing country unit. Let us note y_i the converted daughter yield deviation of animal i, born from sire s and dam d :

$$y_i = b \text{ DYD}$$

The b-factor could be the conversion factor provided by Interbull and derived from the international evaluation with a MACE, or the b-factor of the official conversion formula, if any (national EBV = a + b foreign EBV), or by default a theoretical b-factor estimated by the ratio of within-birth year standard deviations of proofs in the exporting and importing countries.

This information is incorporated into the national evaluation by generating n_i virtual daughters with one performance equal to y_i. The number of virtual daughters is derived from the reliability R_i of i's foreign proof, and from the reliability R_{ip} of i's foreign pedigree index ($R_{ip} = (R_s + R_d)/4$)

$$n_i = n_t - n_p$$

with $n_t = \lambda\, R_i / (1-R_i)$, $n_p = \lambda\, R_{ip} / (1-R_{ip})$, and $\lambda = (4-h^2)/h^2$.

The other parents of these virtual daughters are assumed to be unknown and put into a special group, with effect g. Environmental effects (a herd effect h for instance) may affect these performances. However, as

Interbull Meeting, Prague, September 7-8 1995

$E(y_i) = 0.5\ E(u_i) + 0.5g + h$, coefficients of g and h are always proportional and equations of g and h are linearly dependent (after absorption of other equations). Consequently, the only estimable function of these two effects is $0.5\hat{g} + \hat{h}$. Later on, by convenience, h is assumed to be null and only g is included in the model.

Without any foreign DYD information, the equation of animal i could be written as :

$$\Delta_i\ \hat{u}_i = r_i$$

with Δ_i being the diagonal term of the left-hand side for animal i, $\hat{u}_i$ being i's EBV, and r_i being the right-hand side for animal i, adjusted for i's parents, progeny and mates EBVs, and for all environmental effects affecting i's performances. After incorporating the foreign information, the equations system is expanded to include the equations of breeding values u_d and the permanent environmental effect p_d of the n_i fictive daughters (n_i=3 in this example), and of the group effect g of the unknown parents of the fictive daughters.

$$\begin{bmatrix} \Delta_i + n_i\alpha d_v/4 & -\alpha d_v/2 & -\alpha d_v/2 & -\alpha d_v/2 & 0 & 0 & 0 & n_i\alpha d_v/4 \\ -\alpha d_v/2 & 1+\alpha d_v & 0 & 0 & 1 & 0 & 0 & -\alpha d_v/2 \\ -\alpha d_v/2 & 0 & 1+\alpha d_v & 0 & 0 & 1 & 0 & -\alpha d_v/2 \\ -\alpha d_v/2 & 0 & 0 & 1+\alpha d_v & 0 & 0 & 1 & -\alpha d_v/2 \\ 0 & 1 & 0 & 0 & 1+\gamma & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1+\gamma & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1+\gamma & 0 \\ n_i\alpha d_v/4 & -\alpha d_v/2 & -\alpha d_v/2 & -\alpha d_v/2 & 0 & 0 & 0 & \sum_i n_i\alpha d_v/4 \end{bmatrix} \begin{bmatrix} \hat{u}_i \\ \hat{u}_d \\ \hat{u}_d \\ \hat{u}_d \\ \hat{p}_d \\ \hat{p}_d \\ \hat{p}_d \\ \hat{g} \end{bmatrix} = \begin{bmatrix} r_i \\ y_i \\ y_i \\ y_i \\ y_i \\ y_i \\ y_i \\ 0 \end{bmatrix}$$

with $d_v = 4/3$ because the other parents of virtual daughters are unknown, $\alpha = \dfrac{1-\rho}{h^2}$, $\gamma = \dfrac{1-\rho}{\rho - h^2}$.

After absorbing all virtual daughters equations, the system simplifies to :

$$\begin{bmatrix} \Delta_i + n_i c & n_i c \\ n_i c & \sum_i n_i c \end{bmatrix} \begin{bmatrix} \hat{u}_i \\ \hat{g} \end{bmatrix} = \begin{bmatrix} r_i + 2n_i c y_i \\ 2c \sum_i n_i\ y_i \end{bmatrix} \qquad [1]$$

with $c = (1-\rho)/(4-h^2)$.

Equation [1] shows that DYD information may be incorporated by manipulating the equations system, without explicitly creating virtual daughters equations. Only (at least) one group effect equation per foreign country should be added to the system.

Henderson (1975) presented a very similar approach to compute intraherd breeding values. His goal was to incorporate the national evaluation information for sires in the within herd evaluation. In his approach, the herd and the national evaluation played the role of the national evaluation and the foreign evaluation, respectively, in the present approach. According to our notations, equation for the sire effect in Henderson's paper was $(\Delta_i + n_i c)\hat{u}_i = r_i + 2(n_i + \lambda)c\ ETA_i$, which is identical to i's equation in [1], assuming $(n+\lambda)\ ETA_i = n_i\ DYD_i$, *ie* assuming that i is randomly and independently sampled in the base population of the herd. In our approach, i is assumed to be sampled in another population, which level is accounted for by the group effect.

Interpretation of the group effect $\hat{g}$

As $E(y_i)=0.5(E(u_i) + g)$, the g effect ($-\hat{g}$) represents the breeding value in France of a foreign animal with a zero foreign evaluation. On the other hand, the g effect ($0.5\hat{g}$) can also be interpreted as the country effect on virtual daughters performances, as in Interbull's international evaluation. Therefore, $-\hat{g}$ is a new estimation of the a-factor of the conversion formula. In the present procedure, the dispersion parameters pertaining to foreign data are chosen a priori (b-factor), whereas the location parameters are estimated (a-factor).

The $\hat{g}$ equation could be simply rewritten as $\hat{g} = \sum_i n_i(2y_i - \hat{u}_i)/\sum_i n_i$. This expression suggests that the g effect is estimated by the comparison of virtual daughters performances with national data.

In his proposal, Henderson(1975) assumed that BLUP evaluations of AI sires were "expressed as deviations from the genetic merit of the base population for the herd", but he mentioned this assumption was "not a trivial problem". Including a group or a herd effect solves this problem.

Incorporating EBV information

For females, DYDs are usually not available. Moreover, they do not include own performances information, which may be the major component of a female EBV. Therefore, the superiority y_i of virtual daughters should be derived from the converted EBV. However, to avoid any redondance with EBVs of related animals, this superiority should not include the pedigree information. Let us derive the expression of y_i.

From [1], the equation of animal i could be written as :

$$[\Delta_i + n_i c]\hat{u}_i \quad + [n_i c]\hat{g} \quad = \quad s_i \; + \; 2cn_i y_i \; + \alpha d_i \hat{u}_p \qquad [2]$$

with $d_i = 2$, 4/3, or 1, according to whether i's sire and dam are known or not,

$\hat{u}_p$ being the pedigree breeding value ($\hat{u}_p = 0.5(\hat{u}_s + \hat{u}_d)$)

$s_i = r_i \; - \alpha d_i \hat{u}_p$, *ie* s_i is the right-hand side adjusted for every effects but i's parents EBVs

To derive y_i, let us assume that the animal i has no connection with the national evaluation (*i.e.* no performance and no progeny in the importing country). Under this assumption, s_i=0, $\Delta_i = \alpha\, d_i$, equation [2] simplifies to

$$[\alpha d_i + n_i c]\hat{u}_i \quad + n_i c\,\hat{g} \quad = 2cn_i y_i \; + \alpha d_i \hat{u}_p$$

and after some algebra,

$$2y_i = \hat{g} \; + \; \hat{u}_p \; + \; (1 + \frac{d_i}{n_i}\lambda)(\hat{u}_i - \hat{u}_p)$$

In absence of national information for the foreign animal i, its EBV should be equal to its converted foreign EBV: $\hat{u}_i = u_i^*$. Therefore, $2y_i$ should be set to

$$2y_i = \hat{g} \; + \; u_p^* \; + \; (1 + \frac{d_i}{n_i}\lambda)(u_i^* - u_p^*)$$

As $\hat{g}$ is estimated in the national evaluation, results are invariant to the value guessed for $\hat{g}$, which may be omitted. Consequently, the superiority to be attributed to each virtual daughter is equal to

$$y_i = 0.5\left[u_p^* + \left(1 + \frac{d_i}{n_i}\lambda\right)(u_i^* - u_p^*)\right].$$

As expected, the deregression coefficient increases when the amount of pedigree information in the foreign EBV increases (d_i may vary from 1 to 2) and when the progeny information (n_i) decreases. When u_p^* is unknown, it may be replaced by the mean converted EBV $\bar{u}_i^*$ of the foreign animals without pedigree included in the national evaluation.

Numerical example

Let us consider a foreign animal i, characterized by the following foreign information : i's sire and dam are known and evaluated with a reliability equal to 0.95 and 0.60, respectively, i is evaluated with a reliability equal to 0.6. The heritability is assumed to be equal to 0.25. The number of virtual daughters is

$$n_i = \frac{4 - h^2}{h^2}\left(\frac{R_i}{1 - R_i} - \frac{R_{ip}}{1 - R_{ip}}\right) = 13.01$$

rounded to $n_i = 13$. The deregression coefficient is equal to $\beta = 1 + \frac{d_i}{n_i}\lambda = 3.30$

APPLICATION

This method has been used to include 1843 male and 9,188 female US proofs in the June 1995 French Holstein evaluation, and 92 male US proofs in the Brown Swiss evaluation. Male and female ancestors were considered separately, with two different groups of unknown parents of virtual daughters. The b-factors were those used in the official USA-France conversion formulae, derived from the comparison of full-sibs sampled in France and in the USA (5) in Holstein and from the comparison of daughters of US bulls in both countries in Brown Swiss. Each group effect could be interpreted as an average a-coefficient estimate including every information component : imported semen, French bulls and cows born from imported embryos or imported as live animals. However, no own foreign information was included for young bulls progeny tested in parallel in France and in another country.

The table 1 presents a comparison of the a-coefficient of the official conversion formulae from the USA to France, with minus the estimates of group effects for unknown parents of virtual daughters, for milk, fat, and protein yields. In Brown Swiss, the results were similar, reflecting that the information used in both situations was basically the same. In Holstein, the group effects derived from the US proofs of males were higher than those derived from females proofs for the three traits, and the a-coefficients of the official conversion formulae appeared in an intermediate position. These results could be interpreted by the different sources of information used to estimate the group effects. The group effect derived from male proofs was mainly estimated by the comparison of the superiority of their US and French daughters. As French daughters were born from imported semen, any preferential treatment affecting their performances might have inflated the a-value estimated with US males. On the other hand, the group effect derived from female proofs was estimated by the comparison of the superiority of these females in the US with the superiority of their progeny in France, particularly of their sons used in artificial insemination. These results suggest a small overestimation of bull dams. In spite of the important changes in the US evaluation system in January 1995, all these results could be due to remaining inconsistencies between national genetic evaluation systems.

REFERENCES

1 Banos, G., G.R. Wiggans, and J.A.B. Robinson. 1992. Comparison of methods to rank bulls across countries. J. Dairy Sci. 75: 2560.

2 Banos, G., B. Bonaïti, M. Carabano, J. Claus, P. Leroy, P. Rozzi, J. Philipsson, G. Swanson, and J. Wilmink. 1993. Report of a joint research project between INTERBULL and COPA/COGECA on the feasibility of a simultaneous genetic evaluation of Black-and-White dairy bulls across the European Community countries. Interbull Bull. 9.

3 Bonaïti, B., D. Boichard, A. Barbat, and S. Mattalia. 1993. Problems arising with genetic trend estimation in dairy cattle. Proc. Interbull Meet., Aarhus, Denmark, August 19-20 1993. Interbull Bull. 8.

4 Henderson, C.R. 1975. Use of all relatives in intraherd prediction of breeding values and producing abilities. J. Dairy Sci. 58: 1910-1916.

5 Mattalia, S., and B. Bonaïti. 1993. Use of full-sib families to estimate the a-coefficient of conversion formulae between countries. Proc. Interbull Meet., Aarhus, Denmark, August 19-20 1993. Interbull Bull. 8.

6 Westell, R.A., R.L. Quaas, and L.D. Van Vleck. 1988. Genetic groups in an animal model. J. Dairy Sci. 71: 1310.

Table 1. Comparison of the official a-coefficients from the USA to France, with the group effects estimated from US male ($\hat{g}_m$) and female ($\hat{g}_f$) proofs in Holstein and Brown Swiss

	Holstein			Brown Swiss	
	Official a-value[1]	$-\hat{g}_m$[2]	$-\hat{g}_f$[3]	Official a-value[1]	$-\hat{g}_m$[2]
Milk (kg)	129	435	-55	128	56
Fat (kg)	-6.3	2.0	-26.2	1.6	0.5
Protein (kg)	-1.6	8.2	-10.9	4.7	2.0

[1] 1995 USA-France conversion a-factor, in 1995 male rolling basis

[2] $\hat{g}_m$ is the effect of the group of unknown dams of the virtual daughters of 1843 US Holstein and 92 US Brown Swiss males with a proof included in the French evaluation

[3] $\hat{g}_f$ is the effect of the group of unknown sires of the virtual daughters of 9188 US females with a proof included in the French evaluation