

Impact of Data Structure on Validation of Genetic Trend in Small Population

S. Biffani, A. Samoré, F. Canavesi & P. Boettcher

A.N.A.F.I. Italian Holstein Breeders Association

Via Bergamo, 292 – Cremona, ITALY

*Corresponding author. Tel.: +39.0372.474235; fax: +39.0372.474203

E-mail address: stefanobiffani@anafi.it

Abstract

In 1994 Interbull adapted three methods of validation of international evaluation. The Method 1 is based on comparison of genetic trends estimated using only first lactations (b_1) versus all lactations (b_T) in the routine evaluation. If an over- or under-estimation was detected, age adjustment factors was pointed out as a potential source of bias. Hypothesis that the observed underestimation could be due to the reduced numbers of bulls was made. In order to verify this hypothesis, official data from Italian Holstein genetic evaluation of February 2000 were used to generate random samples containing progressive increasing numbers of bulls (15, 30, 60, 120 and 240 bulls). The genetic trends for milk and protein yields were estimated for the 2760 bulls and for all the generated samples. The results of regression of EBVs on year of birth of the 2760 bulls validated the Italian Holstein evaluation. The samples with a small number of bulls (15) had a greater variation in Δ , ranging from a minimum of – 25,1 kg to a maximum of 45,2 kg for milk and from –0,45 kg to 1,26 kg for protein. These results confirmed the existence of a correlation between apparent bias in estimation of genetic trend and the number of bulls available for international evaluation. These problems could become more evident with the addition of new countries with small populations or new breeds within the current countries, suggesting the development of new methods or standards that take into account the population size.

Keywords: Genetic evaluation; estimated genetic trend; small populations

Introduction

In order to participate in the international evaluation service by the INTERBULL centre, a country must validate its data regarding estimation of the genetic trend, using the three methods originally developed by Bonaiti *et al.* (1994). Method 1 is based on comparison of genetic trends estimated using only first lactations versus all lactations in the routine national evaluations. An important difference (Δ) in the estimated genetic trend would indicate the existence of bias (Bonaiti *et al.*, 1993). According to the INTERBULL Steering Committee, Δ must be less than $.02 \cdot \sigma_g$ ($.01 \cdot \sigma_g$) if breeding values (transmitting abilities) were considered in trend estimation. These fixed thresholds are rather reasonable values and the procedure seems to work correctly with the principal dairy cattle population.

Estimated breeding values (EBV) for production traits of Italian Jersey were calculated for the first time in 2000, using a single trait repeatability animal model according to the procedures described by Wiggans *et al.* (1988). The implementation of a National genetic evaluation procedure was the result of the growing interest of the Italian breeders. Nevertheless, the Italian Jersey population is, from a genetic point of view, the result of independent decisions with regard to importation rather than the result of a national selection program with genetic goals clearly defined. From a numerical point of view, the size of the population is considerably small (6,146 active cows).

The results of the first genetic evaluation (for milk, fat, and protein) for the Italian Jersey breed were submitted in March 2001 to the Interbull Test-

run, jointly with Method 1 results. Fifteen bulls were included in the Method 1 procedure. The calculated Δ s for the three traits were equal or slightly smaller than $.02 \cdot \sigma_g$ and the genetic trend seemed to be slightly under-estimate ($b_1 > b_t$). Even if pre-adjustment factors have been identified as potential source of bias, the obtained results led us to consider the hypothesis that the observed underestimation could be due to the reduced numbers of bulls.

In order to verify this hypothesis, data from a big and previously validated population were used to generate random samples containing progressive increasing numbers of bulls. By doing so it was possible to investigate the existence of a correlation between number of bulls sampled and bias in the estimated genetic trend.

The aim of the present work was to verify to what extent the trend validation procedure was influenced by the sample of bulls available for international evaluation.

Material and Methods

Official files from Italian Holstein genetic evaluation of February 2000 were used to test the effect of the number of bulls included in the international evaluation on the comparison of genetic trends based on primiparous and pluriparous cows.

Two files were available: the first one included the bulls' EBVs for milk and protein yields (47,806 bulls) based on three lactations and the second one included the bulls' EBVs for milk and protein yields (27,647 bulls) resulting from the analyses of primiparous cows only.

The two files were merged and only the proofs of the officially proven bulls born after 1979 and with a minimum of 10 daughters (both primiparous and pluriparous) were considered. The resulting file included 2760 AI Holstein bulls.

The genetic trends for both traits were estimated based on primiparous and pluriparous cows and the difference between the b values was computed.

Successively, the dataset containing the EBVs of the 2760 bulls was sampled to generate sub datasets containing 15, 30, 60, 120 and 240 bulls, using a random number generator of SAS (1998) Twenty

replicates were generated for each sub dataset. A maximum of 200 daughters for each bull was allowed.

The genetic trends for both traits based on primiparous and pluriparous cows were estimated for all the samples, calculating the differences between the b values and computing some descriptive statistics for each group of 20 replicates.

Results were compared with the Δ obtained using all the bulls included in the international evaluation.

Results and Discussion

The results of regression of EBVs on year of birth of the 2760 bulls, the resulting Δ s and the $.02 \cdot \sigma_g$ value for the two traits are presented in Table 1. The yearly genetic trend based on first lactations only (b_1), when compared to trend based on all lactation (b_t), was 98,4% for milk and 97,4 % for protein. The obtained Δ s were less than the $.02 \cdot \sigma_g$ threshold (14,5 kg for milk and 0,4 kg for protein) fixed by Interbull. These results validated the Italian Holstein evaluation. First lactations and later lactations are not the same genetic trait, and a lower trend for first versus later lactations was expected.

In Table 2 (milk) and Table 3 (protein) the results obtained from the analyses of the sub datasets are presented. The results are based on 20 replicates for each sample size. The samples with a small number of bulls (15) had a greater variation in Δ , ranging from a minimum of -25,1 kg to a maximum of 45,2 kg for milk and from -0,45 kg to 1,26 kg for protein. When the sample consisted of only 15 bulls, 40% and 65% of the observed Δ s were greater than the $.02 \cdot \sigma_g$ value for milk and protein, respectively.

In such a situations, similar from the numerical point of view to the one of Italian Jersey, the trends would not be validated by the Interbull thresholds even if the national animal model was correct (Table 1).

In contrast when more than 100 bulls were compared, all the computed Δ s were less than $.02 \cdot \sigma_g$ threshold, both for milk and protein. These results confirmed the existence of a correlation between apparent bias in estimation of genetic trend and the number of bulls available for international evaluation.

From a statistical point of view these results were quite expected, but they have to be taken into account when one deals with validation of genetic trend for small populations. The threshold fixed by INTERBULL is a rather reasonable value for large populations, but for small population it may be too restrictive. Unfortunately, small sets of data create the opportunity for all kinds of problems. The addition of one or few new bulls may have a large impact on trend validation also if the bulls happen to have unusually high or low EBV.

Conclusions

The size of the sample of sires submitted for evaluation influenced the process for validation of trend required by Interbull.

Estimates were consistently within the limits required by the steering committee when the number of bulls available for international evaluation exceed a minimum level (> 60 bulls). With small populations the estimates obtained were highly variable, thus suggesting that the genetic trend was over or under estimated by the evaluation. In reality the national evaluation may not be in these situations.

Pre-adjustment factors have always been considered to be one of the major factors affecting genetic trend validation, as stated by Bonaiti *et al.* (1993, 1994) and by Bagnato *et al.* (1994). This problem was also highlighted in the recent draft of "Interbull Guidelines" (2001). Nevertheless the population size also influences directly the results of trend validation, independently from the definition of pre-adjustment factors.

These problems could become more evident with the addition of new countries with small populations or new breeds within the current countries. Therefore, the development of new methods or standards that take into account the population size may be warranted.

A potential solution, following the example of Guernsey breed, could be the implementation of an International genetic evaluation starting directly from performance records.

References

- Wiggans, G.R., Misztal, I. & Van Vleck, L.D. 1988 Implementation of a n animal model for genetic evaluation of an animal model for genetic evaluation of dairy cattle in the United States. *J. Dairy Sci.* 71 (suppl.2), 54-69.
- Bonaiti, B., Boichard, D., Barbat, A. & Mattalia, S. 1993. Problems arising with genetic trend estimation in dairy cattle. Proc. Interbull Meet. Aarhus, Denmark, August 19-20 1993. *Interbull Bulletin* 8.
- Bonaiti, B., Boichard, D., Barbat, A. & Mattalia, S. 1994. Three methods to validate the estimation of genetic trend in dairy cattle. Proc. Interbull Meet. Ottawa, Ontario, Canada, August 5 – 6 1994. *Interbull Bulletin* 10.
- Bagnato, A., Canavesi, F. & Rozzi, P. 1994. Effect of parity in age adjustment factors in the Italian Holstein Friesian cattle breed. *Proc. 5th WCGALP*. Guelph, Ontario, Canada, August 7 – 12, 1994.
- Interbull Guidelines for National & International Genetic Evaluation Systems of Production traits in Dairy Cattle.* (draft, 2001).
- SAS/STAT User's Guide*, Version 6.12 (1998).

Table 1. Estimated genetic trends in AI Holstein bulls

Trait	b_T (kg)	b_1 (kg)	Δ (kg)	$.02*\sigma_g$ (kg)
Milk yield	131,2 $\pm$ 3,1	129,8 $\pm$ 3,2	1,4	14,5
Protein yield	4,9 $\pm$ 0,08	4,8 $\pm$ 0,08	0,1	0,4

Table 2. Means and relative statistics for the Δ s computed for each sample. Milk yield

$\Delta = b_T - b_1$								
N. samples	N. bulls/sample	Mean (kg)	Max (kg)	Min (kg)	Std Dev (kg)	Std Error (kg)	$.02*\sigma_g$ (kg)	N. samples out of range
20	15	9,5	45,2	-25,1	17,2	3,8	14,5	8
20	30	3,7	22,0	-13,9	9,1	2,0		3
20	60	2,1	17,8	-8,7	6,4	1,4		1
20	120	1,3	9,1	-9,1	4,9	1,1		0
20	240	1,2	6,1	-4,4	3,0	0,7		0

Table 3. Means and relative statistics for the Δ s computed for each sample. Protein yield

$\Delta = b_T - b_1$								
N. samples	N. bulls/sample	Mean (kg)	Max (kg)	Min (kg)	Std Dev (kg)	Std Error (kg)	$.02*\sigma_g$	N. samples out of range
20	15	0,37	1,26	-0,45	0,46	0,10	0,4	13
20	30	0,22	0,81	-0,26	0,29	0,06		4
20	60	0,18	0,73	-0,20	0,20	0,04		3
20	120	0,16	0,41	-0,07	0,14	0,03		1
20	240	0,17	0,31	0,04	0,08	0,02		0