Random vs fixed contemporary group effects on genetic evaluations of the Brazilian Holstein cattle using an autoregressive animal model

D.A. Silva[1], A.A. Silva[1], C.N. Costa[2], F.F. Silva[1], P.S. Lopes[1], S.E.F. Guimarães[1], G. G. Santos[2], G. Thompson[3,4] & J. Carvalheira,[3,4,]*

[1] *Universidade Federal de Viçosa, Departamento de Zootecnia, 36570-000 Viçosa-MG, Brazil*
[2] *Embrapa Gado de Leite, 36.038-330 Juiz de Fora, MG, Brazil*
[3] *Research Center in Biodiversity and Genetic Resources (CIBIO-InBio), Univ. of Porto, 4485-661 Vairão, Portugal*
[4] *Institute of Biomedical Sciences Abel Salazar (ICBAS), Univ. of Porto, Porto, Portugal*
[*] *jgc3@cibio.up.pt (Corresponding Author)*

Summary

Test-day milk yields and somatic cell score (SCS) of Brazilian Holstein cattle were used to compare the effect of random vs fixed contemporary groups (herd-test-date, HTD) using an autoregressive testy-day animal model for genetic evaluations. Fitting the HTD effect as fixed may minimize prediction bias, but requires a minimum number of observations per contemporary group (CG) to simultaneously maximize the effective number of observations and minimize the residual error and prediction error variance. Because most dairy herds in Brazil are characterized by their small size (many CG have less than 3 observations), an alternative is to consider HTD as a random effect especially if non-random associations with sires are considered negligible. A total of 3,428,352 and 2,853,695 testy-day records from the first three lactations of milk yield and somatic cell scores (SCS), respectively, were used for this study. This data was provided by the Brazilian Holstein Cattle Breeders Association and obtained from 1994 to 2016. The additive genetic variances estimated with random CG increased in both traits but also did the estimated phenotypic variances, leaving the heritabilities almost the same. Nevertheless, the annual genetic gains for milk yield (SCS), favored the random CG approach where for males, the yearly progress was 42 kg (-0.034 scores) greater (smaller) than when the fixed CG was fitted. Similarly, for females, the yearly progress was 27 kg (-0.002 scores) greater (smaller) than when the fixed CG was fitted to the model. Steeper genetic trends with random CG indicate that younger animals tended to have higher EBVs giving emphasis to the selection program. These results indicate that using CG as a random effect within the autoregressive animal model should be appropriate for the genetic evaluations of milk yield and SCS for the Brazilian Holstein cattle.

Keywords: *autoregressive animal model, contemporary groups, genetic trends*

Introduction

In an animal model genetic evaluation framework, the minimum number of observations within each contemporary group (CG) is a critical factor to reduce bias on the prediction of breeding values. Fitting CG as fixed or random is a matter of balance between accuracy and bias. Using the random approach may recover some information across CG (Visscher and Goddard, 1993; Strabel et al., 2005) especially when the number of observations in some CG is small, but it will introduce bias to the breeding value predictions if there are non-random associations of animals with effects considered to be fixed (Vasconcelos *et al.*, 2008). If these

non-random associations may be considered negligible, assuming the CG as a random factor will contribute to recover a lot of information that otherwise would be lost. This may be the case of Brazil where most dairy herds are characterized by their small size implying that many of the CG (levels of herd-test-dates, HTD) has less than 3 observations. If these farms are not included in the evaluations, the sire EBV may be biased because of the loss of the information on their daughters, with negative consequences on the genetic progress (Vasconcelos *et al.*, 2008).

Previous studies using test-day records of Holstein cattle in Brazil indicated the AR model (Carvalheira et al., 2002) as a suitable approach to predict breeding values and realize larger genetic gains by selection of production traits (Costa et al., 2014). Therefore, the aim of this study was to compare genetic evaluation for milk yield and somatic cell score (SCS) using an autoregressive testy-day (AR) model considering the effect of CG as random or fixed in the genetic evaluations of Brazilian Holstein cattle.

Material and methods

A total of 3,428,352 and 2,853,695 testy-day (TD) records from the first three lactations of milk yield and SCS, respectively, were provided by the Brazilian Holstein Cattle Breeders Association for the period between 1994 and 2016. The data set was edited according to pre-defined criteria for genetic analysis with AR models (Carvalheira *et al.*, 2002). The pedigree file consisted of 353,681 animals from which, 7912 were bulls and 345,769 were cows. Estimates of the variance components, fixed effects and predicted breeding values were calculated using an AR animal model. The models using the CG as fixed (HTDF) or random (HTDR) can be described as follows:

$$y = X_1\beta_1 + X_2\beta_2 + X_3\beta_3 + Za + Mp + Qt + r.$$

Where, for HTDF, $N(X\beta, V)$ is the vector of TD records, $\beta_{1,2,3}$ are the vectors of fixed effects of HTD, age within herd and DIM within herd and lactation, respectively, a is the vector of random animal additive genetic effects, p is the long-term environmental (LTE) random effect, t is the short-term environmental (STE) random effect and r is the vector of random residual effects fitted with a heterogeneous covariance structure by lactation. X, Z, M and Q are incidence matrices relating observations to fixed and random effects.

For the HTDR, the same definitions apply except for X_1 and β_1 that now refer to the random effect of the CG. A first order autoregressive covariance structure was assumed for HTD (when random), LTE and STE effects. Variance components were estimated using DFREML procedures following assumptions of the AR model (Carvalheira *et al.*, 2002). Convergence criterion was met when the variance of the simplex was less than 10^{-8}. Genetic trends, the Likelihood ratio test (LRT) and rank correlations were used to evaluate the goodness of fit for both models.

Results and discussion

For milk yield, the calculated HTDF likelihood function $[-2\log(L)]$ was smaller than for HTDR, indicating a better fit for the model considering the CG as a fixed effect. For SCS was the opposite, the $-2\log(L)$ was smaller for the random CG approach. For 2 degrees of freedom ($\chi^2_{0.05,2} = 5.9$) the LRT was significant for both cases.

Variance components, autocorrelations and genetic parameters for milk yield and SCS,

are shown in Table 1. The additive genetic variances estimated with random CG increased in the two traits but also did the estimated phenotypic variances, leaving the heritabilities almost the same (Table 1). Costa *et al.* (2014), also using an AR model with fixed CG, analyzed Brazilian Holstein data but restricted the size of the CG to those with 3 or more observations. They found slightly better estimates for heritabilities (0.23 to 0.35 and 0.19 to 0.20 for milk yield and SCS, respectively). The estimated variance associated with the animal non-genetic component was especially important between TD within lactations (STE), revealing a strong permanent environmental effect at this stage. This pattern may be expected from breeds with high genetic potential, producing in high input production systems (Carvalheira *et al.*, 2002).

Table 1. *Estimates of variance components, autocorrelations and genetic parameters for milk yield and somatic cell score (SCS) for Brazilian Holstein cattle, assuming contemporary groups as random (HTDR) or fixed (HTDF) effects in AR models.*

Estimates[1]	HTDF		HTDR	
	Milk ± SE	SCS ± SE	Milk ± SE	SCS ± SE
	10.32 ± 0.113	0.57 ± 0.010	11.03 ± 0.115	0.60 ± 0.012
	<0.001 ± <0.0001	<0.001 ± <0.0001	<0.001 ± <0.0001	<0.001 ± <0.0001
	<0.001 ± <0.0001	<0.001 ± <0.0001	<0.001 ± <0.0001	<0.001 ± <0.0001
	11.74 ± 0.100	1.08 ± 0.019	11.68 ± 0.095	1.08 ± 0.019
	14.49 ± 0.215	1.18 ± 0.014	14.39 ± 0.214	1.18 ± 0.013
	14.96 ± 0.225	1.12 ± 0.012	14.87 ± 0.227	1.12 ± 0.012
	21.31 ± 0.321	1.46 ± 0.021	21.26 ± 0.325	1.46 ± 0.022
	0.69 ± 0.002	0.85 ± 0.002	0.69 ± 0.002	0.85 ± 0.002
	29.38 ± 0.578	1.83 ± 0.036	29.43 ± 0.557	1.83 ± 0.037
	0.78 ± 0.002	0.83 ± 0.001	0.78 ± 0.002	0.82 ± 0.001
	36.26 ± 0.729	2.32 ± 0.044	36.66 ± 0.721	2.31 ± 0.043
	0.76 ± 0.001	0.84 ± 0.002	0.76 ± 0.001	0.84 ± 0.002
	-	-	3.19 ± 0.139	0.27 ± 0.012
	-	-	0.72 ± 0.012	0.63 ± 0.012
	0.24 ± 0.004	0.18 ± 0.002	0.23 ± 0.002	0.18 ± 0.002
	0.19 ± 0.004	0.16 ± 0.002	0.19 ± 0.003	0.15 ± 0.002
	0.17 ± 0.003	0.14 ± 0.002	0.17 ± 0.002	0.14 ± 0.002

[1] : additive genetic variance (kg² for milk and score units² for SCS); : LTE variance; : STE variance; : herd-test-day (HTD) variance; : residual variance; , e : autocorrelations for LTE, STE and HTD; : heritability, where =1, 2 and 3 correspond to first, second and third lactations, respectively.

The influence of random and fixed CG effects on the genetic trends of milk yield and SCS are illustrated in Figure 1. Assuming the CG as random had an enormous impact on the estimated genetic progress of the Brazilian Holstein cattle. A much steeper trend was estimated for the HTDR model in both traits. Annual genetic gains for milk yield were 82.14 kg/yr. compared with 39.95 kg/yr. for cows and 57.96 kg/yr. compared with 30.94 kg/yr. for bulls, respectively. For SCS, annual genetic gains (score units/yr.) were -0.037 *vs.* -0.003 and -0.022 *vs.* -0.020 for cows and bulls, respectively.
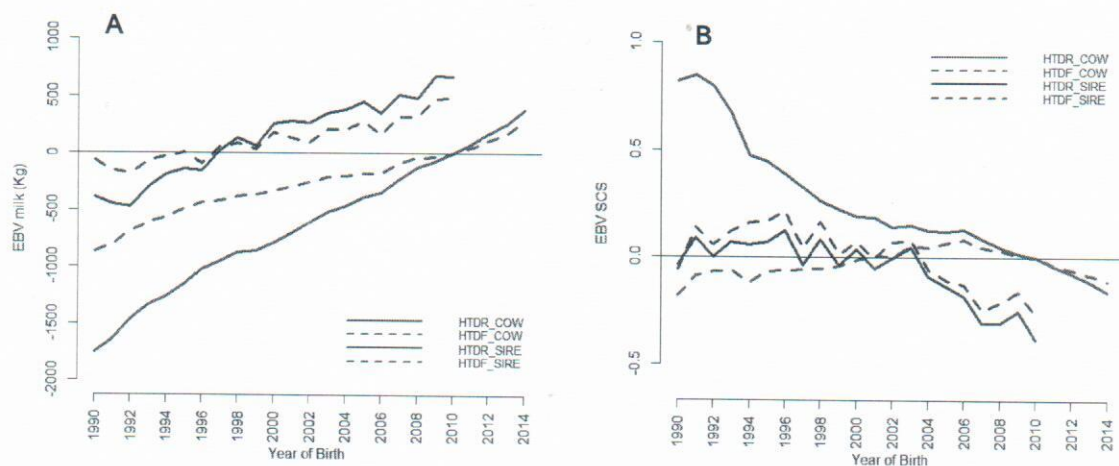
*Figure 1. Annual genetic progress for milk yield (**A**) and somatic cell score (SCS - **B**) for Brazilian Holsteins sires and cows born between 1990 and 2014 considering contemporary group as random (HTDR) or fixed (HTDF).*

Overall, the estimated breeding values and genetic gains per year of birth for the two traits were higher using the HTDR model. Rank correlations between the two evaluations resulted in small deviations on the ranking of animals with the coefficients ranging from 0.90 to 0.93 for milk yield and from 0.67 to 0.87 for SCS. The amount of information recovered by using random CG was considered to have a strong and positive effect on the genetic evaluations suggesting that HTDR model is appropriate in the genetic evaluations of milk yield and SCS of the Brazilian Holstein cattle.

Acknowledgments

List of References

Carvalheira, J., Pollak, E.J., Quaas, R.L., Blake, R.W., 2002. An autoregressive repeatability animal model for test-day records in multiple lactations. J. Dairy Sci. 85, 2040–2045.

Costa, C.N., Cobuci, J.A., Santos, G.G., Thompson, G., Carvalheira, J., 2014. Genetic parameters of test day records in Brazilian Holstein catle using an autoregressive multiple lactation animal model, in: 10th World Congress of Genetics Applied to Livestock Production. Vancouver, Canadá, pp. 5–7.

Strabel, T., J. Szyda, E. Ptak, J. Jamrozik, 2005. Comparison of Random Regression Test-Day Models for Polish Black and White Cattle. J. Dairy Sci. 88, 3688–3699.

Vasconcelos, J., Santos, F., Bagnato, A., Carvalheira, J., 2008. Effects of Clustering Herds with Small-Sized Contemporary Groups in Dairy Cattle Genetic Evaluations. J. Dairy Sci. 91, 377–384.

Visscher, P.M., Goddard, M.E., 1993. Fixed and Random Contemporary Groups. J Dairy Sci. 76, 1444–1454.