Genetic Evaluation of Cow Survival Using a Lactation Random Regression Model

N. Gengler,[*,†] *S. Vanderick,*[†] *P. Mayeres,*[†] *A. Gillon,*[†] *C. Croquet,*[*,†]
[*]*National Fund for Scientific Research, B-1000 Brussels, Belgium*
†Animal Science Unit, Gembloux Agricultural University, B-5030 Gembloux, Belgium

Introduction

Until May 2005 the Walloon region of Belgium was lacking a genetic evaluation system for longevity. There was an obvious consensus that such a system was important. There were different issues that were important. First the new system had to be flexible, robust and easy to implement using existing programs. It had to be able to rely solely on milk recording data. This was one of the most important issues that had to be solved as relatively few animals are under milk recording and therefore cow movement are difficult to track. The system should have been if possible an animal model. A last point that was stressed by breeders was that the model should be able to take into account the whole life of a cow, all her lactations.

Materials and Methods

Choice of model

There is still a certain degree of diversity in the genetic evaluation models for longevity. Roughly one can say that there are 3 groups of models used (INTERBULL, 2005): productive life or lifespan models and survival models where those models can be subdivided into linear and non-linear (survival analysis) models. Current tendency in models used in longevity evaluations is the use of the survival analysis (Ducrocq, 1994). However in our situation we considered an alternative for several reasons. Most experience with survival models showed that their implementation is not always straight forward and extensive fine tuning is necessary. We lacked the needed manpower. Only recently an extension to an approximate animal model was implemented for survival analysis, we considered however an animal model as the best option. The most important reason to choose another system was the fact that we wanted to use optimally the already known survival history of a given cows, allowing to model her survival at a given moment in a given herd amongst other, similar, animals in their respective contemporary groups. This favoured an approach that had similar bases as the currently in Canada used MT model (INTERBULL, 2005) which considers survival in the first three lactations as traits. Australia uses a repeatability model on survival data for cows over successive years (INTERBULL, 2005). A natural extension to this type of models is a random regression model that combines both idea modelling genetic differences across lactation survival with a computational straight forward extension to a repeatability model. Veerkamp *et al.* (2001) showed that this method also combines some of the advantages of a MT approach and the more sophisticated proportional hazards models. The analogies to test-day models allowed also to use similar solving and reliability computation programs which was one of our original aims.

Data

Data was provided from the regular milk recording system. Also available termination codes were too uncertain to be used. As the current population evolved by substituting Dual-Purpose Belgian Blues by Holsteins we exclude for the moment all animals that were not at least 75% Holstein or Red-Holstein. This is only a temporary restriction as we hope that additional research may show ways to include all breeds. Additional data editing consisted in the validation of lactations being inside an acceptable age frame. An, even incomplete, first lactation was required. The observed trait was defined as cow survival in a given lactation. It was coded as 0 (no–survival) if the animal did not calved back, and 1 (survival) if it calved back. If no test for the last herd in the year after the calving year the status of the ongoing lactation is put to a missing record. In total 1,130,533 lactations were available recorded on 392,890 cows. Of these lactation 839,799 had the survival code 1, the other 290,734 the code 0.

Table 1. Statistics on the Holstein data used for Walloon official evaluation for Holsteins in May 2005.

Parity	Lactations	Proportion
1	392890	0.348
2	274596	0.243
3	186588	0.165
4	121808	0.108
5	74496	0.066
6	41691	0.037
7	21585	0.019
8	10022	0.009
9	4315	0.004
10	1611	0.001
11	603	0.001
12	213	0.000
13	74	0.000
14	27	0.000
15	10	0.000
16	3	0.000
17	1	0.000

Table 1 reports details on the Holstein data used for this study. Of the initially present cows less than 1/3 reached the fourth lactation and less than 1/10 reached the tenth lactation.

Models

Definition of random regressions. A major problem in random regression models is that it is not always obvious what type of regressions to use. In test-day models two types of function were tested and used, biological or mathematical (e.g. polynomials). Both have advantages and disadvantages due to the fact that random regressions play a double role. They have to described at the same time the mean and the variance of the underlying biological process. Biological based function are obviously often better for the first purpose, polynomials better for the second. Recently alternative methods were presented as linear splines (e.g., White *et al.,* 1999). In the setting of multi-lactation models another alternative method was recently proposed (Wiggans and Van Raden, 2004). In their method they defined parity differences (**PD**) as the expected a priory change in genetic merit across parities. They defined it with 1 parameter. We applied this idea based on the proportions shown in Table 1. As these proportions are direct function of survival probabilities they should reflect a least partially the expected change in genetic merit across parities. We added a quadratic function of **PD** in order to allow non-linear (quadratic) variation of genetic merit. A constant genetic effect was also introduced.

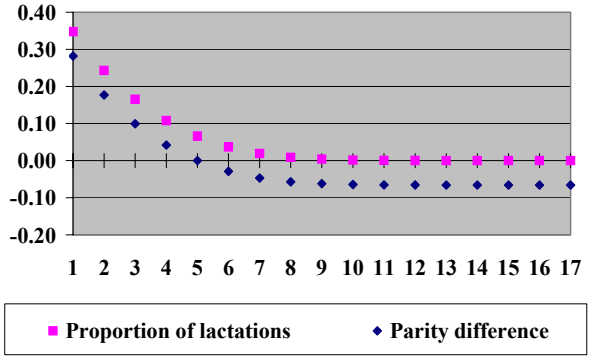

Figure 1. Proportion of lactations and parity differences (PD) used in the random regression model.

The use of constant linear and quadratic functions of PD however did not guarantied correct representation of genetic variances. Therefore additional steps were taken to adjust regressions during the variance components estimation procedures.

Table 2. Final regressions and weights used.

	Regressions			
Parity	1	2	3	Weight
1	0.415	0.405	0.145	2.798
2	0.186	0.711	0.071	1.833
3	0.497	0.589	0.023	1.682
4	0.807	0.238	0.003	1.413
5	1.000	0.000	0.000	1.000
6	0.992	-0.125	0.001	1.000
7	0.984	-0.180	0.002	1.000
8	0.978	-0.207	0.003	1.000
9	0.976	-0.219	0.003	1.000
10	0.975	-0.224	0.004	1.000
11	0.974	-0.226	0.004	1.000
12	0.974	-0.227	0.004	1.000
13	0.974	-0.227	0.004	1.000
14	0.974	-0.227	0.004	1.000
15	0.974	-0.227	0.004	1.000
16	0.974	-0.228	0.004	1.000
17	0.974	-0.228	0.004	1.000

Estimation of genetic parameters. Table 2 gives final uncorrelated regressions and weights used in the computations. A major challenge for this work was the estimation of variance components. They were developed from various sources. Initial estimations of correlations among first three parities were taken from literature (e.g., Jarirath *et al.,* 1998). Based on a subset of

data correlations and variances across the first five parities were obtained. Heritability of survival in a given parity was put to 0.03 for all parities. Genetic (co)variances among initial regressions were obtained by backsolving. Regressions were transformed and standardized by adjusting regressors and therefore variances towards the variance of the fifth parity. Residuals were assumed uncorrelated and residuals variances standardised towards the fifth lactation using weights. Parities after the fifth were considered having the same variances as the fifth. Table 3 gives details on the variances and correlations used. The most interesting are the results for the second lactation. This lactation behaves differently from the others. At the present time we are unsure if this is an artefact or at least partially reality. We will in the next months proceed with additional variance component estimations to clarify this. Given its still rather high correlations with first and third lactation survival we do not expect major ranking changes if correlations are considered being higher than those used in the present computations.

Table 3. Genetic correlations, genetic variances (on diagonal) and residual variances (residual correlations being 0) across selected lactations (1 to 7, 10 and 17).

Parity	1	2	3	4	5	6	7	10	17
					Genetic				
1	0.00232	0.850	0.972	0.858	0.694	0.604	0.561	0.525	0.523
2	0.850	0.00354	0.900	0.514	0.251	0.129	0.074	0.029	0.026
3	0.872	0.900	0.00386	0.835	0.645	0.545	0.497	0.458	0.455
4	0.858	0.514	0.835	0.00496	0.959	0.917	0.893	0.871	0.870
5	0.694	0.251	0.645	0.959	0.00650	0.992	0.984	0.975	0.974
6	0.604	0.129	0.545	0.917	0.992	0.00650	0.998	0.995	0.995
7	0.561	0.074	0.497	0.893	0.984	0.998	0.00650	0.999	0.999
10	0.525	0.029	0.458	0.871	0.975	0.995	0.999	0.00650	0.999
17	0.523	0.026	0.455	0.870	0.974	0.995	0.999	0.999	0.00650
					Residual				
	0.0751	0.115	0.125	0.149	0.210	0.210	0.210	0.210	0.210

Genetic evaluation model. The general evaluation model was voluntarily kept extremely simple. It can be written as :

$$\mathbf{y = Xh + Wb + ZQa + e.}$$

where **y** is a vector, **h** is a vector of fixed contemporary effects based of herd x quota-year (1st april to 31st of march) of calving x parity-group (1, 2, 3, 4 and +), **b** is a vector of fixed effects of birth-year x parity-group (1, 2, 3, 4 and +), **a** is a vector of the three random regression effects per animal, **e** is vector of the random residual effects, **X**, **W** **Z** are incidence matrices linking observations to the effects and **Q** is a matrix of regressors. The trend correction effect (**b**) plays an important role because it allowed us to take a least partially into account the trend due to selection on correlated traits. We prefer doing this for the moment to regressions on phenotypes because it is still unclear what the reasons for voluntary culling in our population are.

The reported trait was defined as the sum of all genetic effects times its associated regression coefficients up to the fifth parity included. Several reasons exist for this choice. First five lactations is a reasonable longevity objective, similar as 305 days is a reasonable lactation length objective. Summing these values is conceptual identical to an equally weighted sum of lactation survival solutions from a MT model. More complicated weighting methods could have been used (e. g., Jairath *et al.,* 1998), however this simpler method is easier too explain. Also using equal weights is more closely related to livespan or productive live type values that are defined over similar windows. The genetic base was defined by putting the mean of all cows with records born in 2000 to 3. This value was chosen because of its simplicity and parallelism to our SCS mean.

Results and Discussion

Table 4. Means and standard deviations of EBV for cows with records and their sires.

Animals	Number	Mean	SD
Cows	392 890	2.72	0.27
Sires of cows	10 237	2.70	0.33

Table 4 gives the means and standard deviations of EBV for cows with records and their sires. The mean value in the whole cow population is close to the expected value of the number of lactations which is around 2.6.

Figure 2 give the genetic trend in the cow population since 1992. The raw trend was slightly positive as expected because of the trait definition. However correcting using a similar strategy as Wiggans and VanRaden (1995) using phenotypic regression coefficients for production (milk, fat protein called V€L) and all indexes: production (V€L), type (V€T) and SCS (V€F) was not able to flatten totally this trend. Additional investigations will be done on this topic.

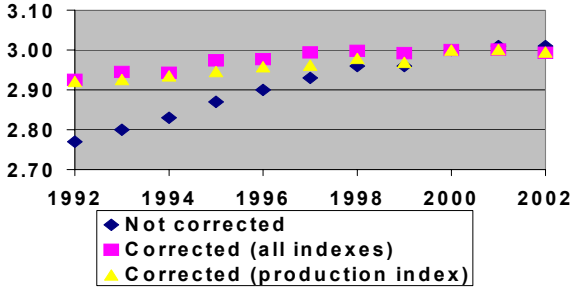

Figure 2. Raw genetic trend, genetic trend corrected for milk index (V€L) and genetic trend corrected for all indexes.

A total of 436 sires were used by INTERBULL in the routine run. Table 5 show statistic comparing the correlations of our evaluation with the other INTERBULL populations. In general our population behaved well and had reasonable correlations with most populations.

Table 5. Mean (of the Walloon Region), overall mean of all populations, minimum and maximum correlations used in the May 2005 routine-run.

Belgium (WR)	Overall	Minimum	Maximum
0.69	0.63	0.33	0.86

Table 6. Correlations among EBVs for longevity and global and partial idexes for INTERBULL 05/2005 sires ≥ 0.60 reliability for longevity.

Global (V€G)	Milk (V€L)	Type (V€T)	SCS (V€F)
0.16	0.02	0.20	-0.36
Development (V€C)	**Feet & Legs (V€M)**	**Udder (V€P)**	
-0.13	0.04	0.31	

Table 6 show the correlations we observed for the 8391 sires with a reliability of 0.60 and more. Interesting is that the trend we noticed for the cows disappears (results not shown). The observed correlations reflect very closely what one might expect. However the results we get back could be considered another trait. Therefore we are still wondering how to harmonize our results for cows and the results for bulls provided by INTERBULL especially considering Dekkers (1993).

Acknowledgments

Nicolas Gengler, who is Research Associate and Coraline Croquet who is Research Fellow of the National Fund for Scientific Research (Brussels, Belgium), acknowledge their support. Additional support was provided through Grant 2.4507.02 F (2) of the National Fund for Scientific Research. The authors gratefully acknowledged the support of the Walloon Breeding Association (AWE) and the Walloon Regional Ministry of Agriculture (MRW-DGA, especially projects : RW1009, D31-1039).

References

Ducrocq, V. 1994. Statistical analysis of length of productive life of dairy cows of the Normande breed. *J. Dairy Sci. 77,* 855-866.

INTERBULL. 2005. Description of National Genetic Evaluation Systems for dairy cattle traits as applied in different Interbull member countries. http://www-interbull.slu.se/national_ges_info2/framesida-ges.htm (Accessed May 31, 2005).

Jairath, L., Dekkers, J.C.M., Schaeffer, L.R., Liu, Z., Burnside, E.B. & Kolstad, B. 1998. *J. Dairy Sci. 81,* 550-562.

Veerkamp, R.F., Brotherstone, S., Engel, B. & Meuwissen, T.H.E. 2001. Analysis of censored survival data using regression models. *Animal Sci. 72,* 1-10.

White, I.M.S., Thompson, R. & Brotherstone, S. 1999. Genetic and environmental smoothing of lactation curves with cubic splines. *J. Dairy Sci. 82,* 632-638.

Wiggans, G.R. & VanRaden, P.M. 1995. Productive life evaluations: calculation, accuracy, and economic value. *J. Dairy Sci. 78,* 631-638.

Wiggans, G.R. & VanRaden, P.M. 2004. Accounting for differences in rate of maturity in yield evaluations. *J. Dairy Sci. 87 (Suppl. 1),* 412 (abstr. 737).

Dekkers, J.C.M. 1993. Theoretical basis for genetic parameters of herd life and effects on response to selection. *J. Dairy Sci. 76,* 1433-1443.