

IMPLEMENTATION OF A ROUTINE BREEDING VALUE EVALUATION FOR LONGEVITY OF DAIRY COWS USING SURVIVAL ANALYSIS TECHNIQUES

V. Ducrocq¹ and J. Sölkner²

¹Station de Génétique Quantitative et Appliquée, INRA, F-78352 Jouy-en-Josas, France

²Dept. of Livest. Sci., U. of Agriculture, Gregor-Mendel-Str. 33, A-1180 Vienna, Austria

SUMMARY

A genetic evaluation of dairy sires based on survival analysis of the longevity of their daughters was implemented in France and in Austria. The model allows the inclusion of censored records and a correction of length of productive life for changes in culling policies over time and for voluntary culling due to low milk production. It is applicable to very large datasets.

Keywords : dairy cattle, genetic evaluation, genetic parameters, longevity.

INTRODUCTION

For decades, selection in dairy cattle has dealt almost exclusively with production traits. Now, functional traits are receiving increasing attention because of their impact on costs. Among such traits, functional longevity, defined as the ability to delay involuntary culling, summarizes the effect of all functional traits on culling: when at least one of them is defective, length of productive life (LPL, measured as the number of days between first calving and culling) is reduced. Short LPLs are associated with increased replacement costs and reduced possibilities of selection on other traits.

LPLs are also influenced by production traits, through voluntary culling of cows with insufficient milk yield. In order to supply breeders with extra information reflecting functional longevity independently from dairy traits, LPLs must be corrected for the phenotypic effect of production.

For some cows, e.g., those still alive, only a lower bound of the exact LPL is available. This is known as censoring. In the past, the need to include censored records in the analysis has been avoided by using indirect longevity indicators such as whether the cow is still alive at a given age or lactation (Schaeffer and Burnside, 1974), or by replacing censored records by projected LPLs based on currently available information (van Raden and Klaaskate, 1993). The former strategy uses only partial information. With the latter, the amount of information available to predict completed LPL records is extremely limited ($R^2 = 0.10$ at 48 mo of age). An alternative is to use both censored and uncensored records in a survival analysis (Cox, 1972). Such an approach relies on the concept of hazard rate, the limiting probability of being culled among animals still alive. The hazard rate can be modelled for all records, whether censored or not. In most cases, the hazard rate is described as the product of a baseline hazard rate, representing the ageing of the population and an exponential function of covariates. In proportional hazards models, the baseline hazard function is often left arbitrary, leading to 'Cox models' (Cox, 1972). Such models including random (genetic)

effects were introduced in animal breeding by Smith and Quaas (1984) and have been used in routine genetic evaluations for LPL in Austria since 1995 (Egger-Danner, 1993).

Another peculiarity of survival data is the fact that the explanatory variables often vary with time: culling policies in a herd change with year, season, quota status, variation in herd size, etc... To fit a single herd (or herd-year-season) effect does not reflect the influence of these variations on LPL adequately. Current milk production on which voluntary culling is based also varies with time. If milk yield is to be included in the model to approximate functional longevity, it has to be time-dependent. Proportional hazards models have been extended to incorporate time-dependent covariates (Kalbfleisch and Prentice, 1980) and applied to the analysis of LPL (Ducrocq, 1987). In fact, it has been shown that the arbitrary baseline hazard function in the proportional hazard model can be advantageously replaced by a parametric (Weibull) hazard function (Ducrocq, 1987). This result, coupled with the extension of mixed survival models to include relationships between sires (Ducrocq and Casella, 1996) and with the availability of the adequate software (Ducrocq and Sölkner, 1994, 1998) offers the possibility of national evaluations of dairy bulls based on the LPL of their daughters. Such evaluations were implemented in France and in Austria in June 1997.

MATERIAL AND METHODS

Implementation in France

The datasets analyzed included all cows milk recorded between December 1, 1984 and May 15, 1997. For cows with first calving before December 1984, only the part corresponding to lactations started after that date was used («left-truncated» records). Records of cows with last test date after January 1, 1997, changing herds during their productive life or in herds disappearing from the files were considered as censored.

The model used for both the variance components estimation and the genetic evaluation is :

$$\lambda(t) = \lambda_0(t) * \exp\{y_i(\tau) + l_j(t, t') + m_k(t) + f_n(t) + p_o(t) + v_q(\tau) + h_r(\tau) + s_u\}$$

where $\lambda(t)$ is the hazard of a cow, t days after her first calving and :

- $\lambda_0(t) = \rho(\kappa t)^{\rho-1} = \rho t^{\rho-1} e^{\rho \log \kappa}$ is the Weibull baseline hazard function with parameters ρ and κ (here, $\rho \log \kappa$ plays the role of a grand mean);
- $y_i(\tau)$ is the time-dependent year-season effect at calendar date τ , with changes on March 1 and December 1 each year;
- $l_j(t, t')$ is the time-dependent combined effect of lactation number, t days after first calving (lactations 1, 2, 3, 4, 5, 6+) and of stage of lactation, t' days after the current calving (with changes at $t'=30, 60, 150, 240$ and date when dried) ;
- $m_k(t), f_n(t), p_o(t)$ are time-dependent classes of within herd-year deviations for milk yield (10 classes), fat content (5) and protein content (5) with changes at each new calving date ;
- $v_q(\tau)$ is the time-dependent combined effect of 4 herd size classes and (up to) 5 variation in herd size classes with changes on March 1 and December 1 each year;
- $h_r(\tau)$ is the time-dependent random herd-year-season effect at τ with changes on March 1 and December 1 each year; its distribution is assumed to be log-gamma(γ, γ).

- s_u is the random effect of the sire of the cow. Sire effects $\mathbf{s} = \{s_u\}$ follow a multivariate normal distribution with variance-covariance matrix $\mathbf{A} \sigma_s^2$.

Preliminary analyses lead to the use of $\rho = 2$ and $\gamma = 4$. All computations were done using « The Survival Kit, V3.0 » (Ducrocq and Sölkner, 1998). The genetic variance of sire effects was estimated as the mode of the (approximate) marginal posterior distribution of σ_s^2 , using a procedure described in Ducrocq and Casella (1996) and the sires' EBVs were obtained as a by-product at convergence. For the Holstein breed, a subset was used for the estimation of σ_s^2 and the genetic evaluation on the full data set was performed separately.

Implementation in Austria

In Austria, the average herd size is very small (12 cows per herd under milk recording, considerably less for some breeds). The use of herd-year-season effects under such circumstances may be problematic. Therefore, alternative ways of dealing with the management strategies of farmers with regard to culling were investigated. Pinzgauer, a dual purpose breed, was taken as an example. The model was different from the French model in the following ways: $v_q(\tau)$, i.e., variation of herd size was excluded (herds too small); $h_r(\tau)$ was replaced by a fixed time-dependent effect of average age of herdmates of a cow (9 classes); for within herd-year deviations of milk yield, fat content and protein content 9 classes were defined each. The dataset analyzed included all cows milk recorded between January 1, 1984 and January 31, 1997. In addition to cows still alive, cows sold for dairy purpose to non-breeders (a conventional practice in this breed) were also treated as censored.

RESULTS AND DISCUSSION

Results in France

Estimates of the sire variance are given in table 1 for the three main breeds in France. The three data files were of similar size and the estimation took about 16 CPU hours on average, on an IBM Risc 6000-AIX41 computer. A formal definition of heritability can be made on the logarithmic scale of LPL (Ducrocq and Casella, 1996). Using a Taylor series expansion of $\log(\text{LPL})$ around its mean, it is possible to get an approximate heritability on the original scale (see table 1). These are higher than in the literature but correspond to *completed* LPL records: the reliability of the sires' EBV depends on the number of *uncensored* daughters, not on the total number of daughters for each sire.

The first genetic evaluation for the Holstein breed took about 38 CPU hours (150 likelihood evaluations, starting from 0). EBVs were published in genetic standard deviations, after a change of sign (positive EBVs correspond to negative \hat{s}_u , i.e., to lower culling rate). Genetic trends were favorable for the three breeds. Within-year correlations between LPL and dairy traits EBVs were still favorable (for the Holstein breed: 0.16 with $\text{INEL} = \text{Protein kg} + 3 \text{ Protein } \%$). Other favorable correlations were found with udder EBV (in particular with udder depth EBV), somatic cell count EBV and milking ease EBV.

Table 1. Datasets characteristics and results for the three main breeds in France

	Montbéliarde ^{1,2}	Normande ^{1,2}	Holstein ¹	Holstein ²
Records (total)	668074	762910	629716	6399885
Censored (%)	36.0	31.4	36.2	32.0
Left truncated (%)	12.0	11.0	8.9	12.7
Sires	3354	4175	2937	17777
Sire variance (\pm std) ³	0.047 \pm 0.002	0.039 \pm 0.002	0.053 \pm 0.002	
h ² (% , original scale)	19.3	16.1	21.7	

¹ Estimation of genetic parameters² Genetic evaluation³ Mean \pm standard deviation of the approximate marginal posterior density

Results in Austria

The total number of records for the Pinzgauer breed was 48242, offspring of 1013 sires. The total number of sires in the pedigree was 1553. The proportion of censored records was 61%, reflecting the transfer of cows in milk from breeders to non-breeders. The estimated sire variance was 0.047, the same as found in the French Montbéliarde breed. The time-dependent effect of average age of herd mates showed to be important. Cows who were contemporaries of herdmates in the lowest age-class were 3.57 times more susceptible to culling than cows with herdmates in the highest one. Correlations between LPL and EBVs for milk were .27 for milk yield, correlations with EBVs for fertility and calving ease were favorable and in the range of .15 to .2. The genetic trend for LPL was close to zero. The correlation between the EBVs for LPL estimated with this model and the EBVs currently published using a Cox model with time-independent covariates only was .85.

CONCLUSION

These applications illustrate that a proper statistical analysis of LPL, including censored and left-truncated records and taking into account the fact that culling policies and milk production vary with time is feasible, even on extremely large data sets.

REFERENCES

- Cox, D.R. (1972) *J. R. Stat. Soc., Series B*, **34** : 187-220.
 Ducrocq, V. (1987) PhD Thesis, Cornell Univ., Ithaca, New-York, USA.
 Ducrocq, V., Casella, G. (1996) *Genet. Sel. Evol.*, **28**: 505-529.
 Ducrocq, V., Sölkner J. (1994) In 'Proc. 5th World Cong. on Genet. Appl. To Livest. Prod.', Vol 22, p.51, Univ. of Guelph.
 Ducrocq, V., Sölkner J. (1998) In 'Proc. 6th World Cong. on Genet. Appl. To Livest. Prod.'.
 Egger-Danner, C. (1993) PhD Thesis, Univ. of Agriculture, Vienna, Austria.
 Kalbfleisch, J.D. and Prentice, R.L. (1989) 'The Statistical Analysis of Failure Time Data' John Wiley and Sons, New-York, USA.
 Schaeffer, L.R. and Burnside E.B. (1974) *J. Dairy Sci.*, **57** : 1394-1400.
 Smith, S.P. and Quaas, R.L. (1984) *J. Dairy Sci.*, **67** : 2999-3007.
 Van Raden, P.M and Klaaskate, E.J.H. (1993) *J. Dairy Sci.*, **76** : 2758-2764.