Genetic Evaluation for Length of Productive Life of Holstein Cattle in Germany

Erik Pasman and Friedrich Reinhardt
Vereinigte Informationssysteme Tierhaltung w.V. (VIT),
Heideweg 1,D- 27283 Verden, Germany

Abstract

Since 1996, genetic evaluation of length of productive life of Black-and-White and Red-and-White Holstein dairy cows is carried out for West Germany. The SURVIVAL KIT package was applied for the first time in 1998 on data of all German HF-cows (Black & White and Red & White). Length of productive life is modelled with a Weibull model considering year-season within herd, lactation number * stage, relative milk yield within herd * year-season, relative change of herd size within year and a combined random genetic effect of sire and maternal grandsire of the cow. Sire solutions are standardized to a mean of 100 and a genetic standard deviation of 12, relative risks and an approximation of the breeding value expressed in days are also published. Solutions of fixed effects were plausible but averages of sire solutions differed among regions. Therefore a test-run was made in which interactions of region with lactation number * stage and with relative milk yield within herd * year-season were modelled. Solutions of the fixed effects showed that relative milk yield within herd * year-season should be considered in interaction with region.

---

1. Introduction

In Germany, all records of Black-and-White and Red-and-White cows calved from 1980 onwards are available for genetic evaluation of functional length-of-productive life (LPL) by survival analysis. Length of productive life is defined as the time from first calving to culling. In 1996 and 1997, genetic evaluation for the Holstein population of West-Germany was carried out with a COX model with stratification to herd-year, relative milk yield and relative percentage fat + protein as regression factors, and a sire effect (Egger-Danner, 1993). Using the SURVIVAL KIT V3.0 for genetic evaluation (Ducrocq and Sölkner, 1994 & 1998a), it is possible to model other effects on LPL too, even when they are time-dependent.

2. Material and methods

2.1. Data

Records of Black-and-White and Red-and-White cows in the milk recording program calved from 1980 onwards are available for genetic evaluation of length of productive life through survival analysis. Length of Productive-Life is defined as the period from first calving to culling. Records of cows still present in the herd or sold for dairy purposes are right-censored. Records of cows in West Germany are left truncated at 1st January 1985 and records of cows in East Germany at 1st January 1991. The left-truncation dates were set to avoid the period of implementation of the milk qouta system (West-Germany) and the period up to re-unification (East-Germany) being considered in the evaluation. Records of daughters of East-German non-Holstein sires are ignored. The reason of exclusion of East-German data till 1991 and daughters of East-German non-Holstein sires is that these cause biases in the estimated breeding values. Cows culled before 46 days in the 1st lactation are excluded because an accurate projection of the 305-day milk yield for these cows was not possible. Age of calving is restricted between 610 and 1320 days for the 1st lactation and between 1000 and 2135 days for the 2nd lactation. Cows with a last lactation of more than 750 days are excluded from the analysis. Currently, genetic evaluation for LPL is carried out once a year for publication in August. The size of the data set used for evaluation in 1998 is described in table 1.

Table 1: ***No. cows, herds and sires in the 1998 evaluation of LPL***

no. cows	7,903,285
of which left-truncated	1,015,134
of which right-censored	2,857,392
no. herds	82,716
no. bulls (occurring as sire and/or mgs)	170,227

2.2. Model

Survival analyses were carried out with the SURVIVAL-KIT V3.0 (Ducrocq and Sölkner, 1994 & 1998a) using a Weibull model. The distribution of the observed random variable LPL is described using the hazard function

$$\lambda(t,z)_j = \lambda_0(t) * e^{(x'(t)b+z'u)}$$

where:

- $\lambda(t,z)_j$: relative risk of culling at time (t) from animal (j) under occurrence of the factors considered in the vectors x(t) and z
- $\lambda_0(t)$: base hazard function
- b: solution vector for all environmental effects considered in x(t) and
- u: solution vector for all genetic effects

The following information is taken into account in the model:

Random time-dependent effect assumed to follow a log-gamma distribution, which is integrated out (Ducrocq and Casella, 1996, Ducrocq and Sölkner, 1998b):
- herd-year-season effect with changes on 1st May, 1st November and 1st January.

Fixed time-dependent effects as class variables in (x'(t)b):
- parity * stage-of-lactation with changes at 60, 180, 270 and 360 days after calving or date when dried (lactations 1, 2, 3, 4, 5, >5)
- relative fat and protein yield within herd * year - season; fat: protein weighted 1:4 as in the milk production index; changes at the first three calving dates (11 yield classes)
- relative change of herd size from 1st January to 31st December (21 classes)

Random genetic effect in (z'u):
- the combined genetic effect of a bull as sire and maternal grand-sire (weight of sire information : mgs-information 2:1).

The breeding value of a bull is two times the genetic effect based on his daughters and maternal grand-daughters. The genetic parameters presented for French Holstein cattle by Ducrocq and Sölkner (1998b) were used in the analyses.

2.3. Presentation and publication of proofs

Breeding values of bulls for LPL are estimated once per year and are published in the form of lists and through direct data access as relative breeding value (RZN), which is a standardized sire solution. The relative risk (the exponential function of the breeding value on the original scale) is only available in direct data access. The approximate transformation (see further down for transformation method) in days-LPL is for interpretation purposes published in lists and direct data access. The base population for the relative breeding values consists of the sires born from 1988 to 1990 with a minimum reliability of the breeding value of 50%. Estimated breeding values of the sires in the base population are standardized to the mean of 100 and a genetic standard deviation (of the true breeding values) of 12. The breeding values have to be inverted to obtain desirable breeding values above 100. Reliabilities are calculated as the approximate percentage of determination (r_{gi}^2) between real breeding values and estimated breeding values. Censored records are weighted with the square of the genetic correlation between LPL at the moment of censoring and actually completed LPL. These correlations were estimated by Egger-Danner (1993).

In figure 1 expected survival curves of average daughters of sires with different RZN-values (88, 100, 112) are presented. The number of daughters starting their productive life with a first calving is put to 100%. For the approximate derivation of the survival curves and the breeding values in days-LPL, an average culling rate in the first lactation for all cows was assumed to be 20%.
The LPL achieved by 50% of the daughters is shown by a vertical line.

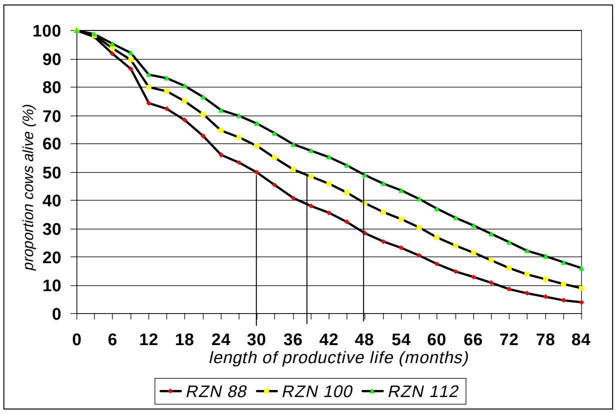

Figure 1: Expected survival curves for different RZN (average culling rate 1st lactation = 20%)

The most important parameters derived from these survival curves are shown in table 2. These show that there is a considerable variation in the genetic transmission of LPL. It is also shown that breeding values for relative risk and days LPL have a non-linear relationship with LPL.

Table 2: Average length of productive life (years / days) for different RZN (average culling rate in 1st lactation in population = 20 %)

RZN	EBV - culling risk relative scale	EBV - LPL years (days)	time 50% daughters culled years (days)
88	1.651	-1.16 (-423)	2.50 (911)
100	1.000	0,00 (0)	3.08 (1124)
112	0.605	+1.42 (+518)	3.92 (1431)

3. Results from the initial model

The data for the analysis consisted of 128,431,407 elementary records. The evaluation was carried out on a HP-K420 computer (PA7xxx processor). It took ca. 8 hours to write the temporary file. Iteration took slightly over 30 CPU-minutes per round and was stopped after 300 rounds. The system did not converge to the desired criterion but the convergence criterion did not decrease further. Averages and standard deviations of the breeding values of the sires in the base population (birth-years 1988 - 90, n=1834) are presented in table 3.

Table 3: Means and standard deviation of breeding values of bulls forming the base (born 1988-90, n=1834)

	mean	standard deviation
EBV - relative culling risk	1.088	0.461
EBV - LPL (days)	35	367
RZN	100.0	9.9

Means of breeding values for relative culling risk and days LPL are not equal to their expectations according to the base definition (1 and 0 respectively), although the average of RZN is 100. This is due to the non-linear relationship of the breeding values for relative culling risk and days LPL with RZN.

Differences among year-season effects were small. The risk of culling is relatively high in the first lactation and from 270 to 360 days after calving. The culling risk more than 360 days after calving is extremely low indicating that nearly all cows in this class will calve again. Seasonal differences in culling for low milk yield were found in 1990 and 1991, but not in 1997. Generally, culling for low milk yield was more severe in 1997. Risk of culling due to change in herd size was as expected considerably higher in shrinking herds. The risk of culling was stable in classes of herds expanding 20% or more.

Considerable differences among regions in average RZN were found. These differences increased with birth year. These differences are expected to be caused by regionally different culling practices, because average herd structure and management in Germany differ by region. Furthermore, results of previous evaluations showed that none of the AI-Stations selected intensively for LPL.

4. Changes in the model

Two test runs were made to investigate regional effects. In the first run, region was included in the model in interaction with lactation number * stage and in interaction with relative yield * year-season. Yields below 75% and above 125% of the herd average were combined, leading to the number of yield classes being reduced to 7. The reason for this is that the occupation of the extreme classes was poor when relative yield * year-season was taken in interaction with region. Breed of the cow also was included in the model, to avoid regional differences due to different breed composition being adjusted for by including region in interactions. Sire solutions were adjusted for breed of their daughters before standardisation to obtain comparable sire solutions across breeds. For instance, half of the breed solution for Black-and-White and half the breed solution of Red-and-White are added to the solution of a sire with an equal amount of information from daughters in both breeds.

The estimation of regional effects appeared not to be correct. Estimated effects for lactation number * stage were for region N substantially lower than for other regions. A conflicting result is that in the same evaluation estimates for relative yield * year season for region N were substantially higher than for the other regions. As solutions for sires from region N were on average clearly higher than those of sires from elsewhere, it was suspected that the inclusion of region in both interactions also leads to biases in the estimation of sire effects.

In the second test-run, region was only included in the model in interaction with relative yield * year-season. The reason for this is that the interaction of region with relative yield * year-season appeared to be more important than the interaction of region with lactation number * stage. The evaluation was carried out on a HP-K460 computer (PA8000 processor). It took ca. 2.5 hours to write the temporary file. Iteration took about 17 CPU-minutes per round. Convergence was reached after 147 rounds.

Results of the second test-run showed differences between regions in the level of culling. These differences varied considerably with relative yield * year-season (see figure 2).

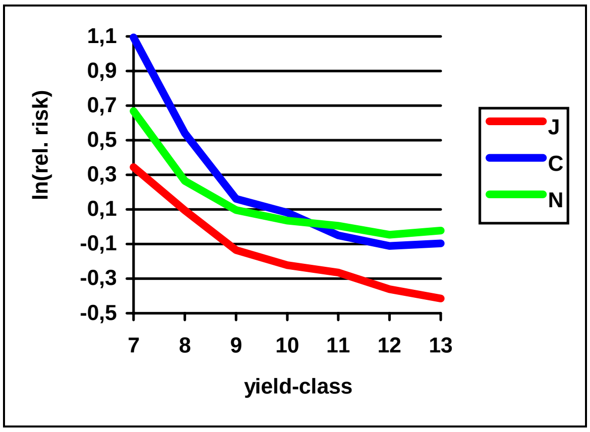

Figure 2: Solutions for relative yield classes in November/December 1997 for regions C, J and N

Standard deviations of estimated breeding values on the original scale dropped from 0.502 to 0.185. The standard deviation of relative risks and EBV for days-LPL dropped to 0.156 and 133 respectively. The standard deviation of RZN did not change because of standardization. Correlations according to birth-year of sire-solutions of the routine evaluation with those of this test-run were around 0.80 for sires born before 1982 and gradually increased to over 0.95 for sires born from 1990. Differences in RZN among regions were smaller than in the official 1998-evaluation. Trends in RZN were similar for all regions.

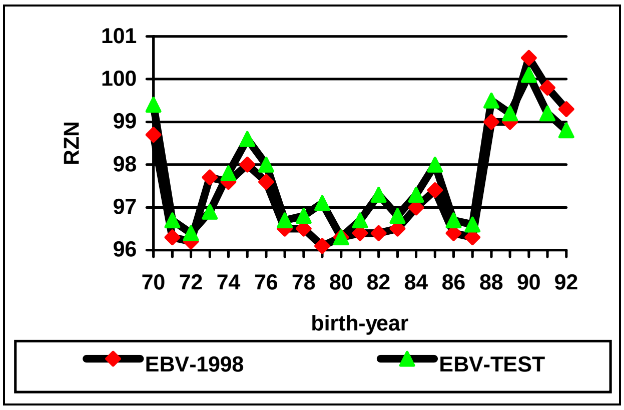

Figure 3: ***Genetic trends for LPL on the RZN-scale (min. rel. = 50%)***

Trends in average breeding values estimated in the routine evaluation and in the second test run were approximately equal (see figure 3). The reason for the low averages for 1986 and 1987 and the high averages after 1987 in both runs is not clear. Possibly, these are caused by a biased projection of LPL of censored cows. Trends in relative risk and LPL-in-days were similar but less pronounced after the test-run. This is due to the smaller standard deviation of the estimated breeding values in the test-run.

5. Discussion

The development of the SURVIVAL-KIT by Ducrocq and Sölkner (1994 & 1998a) offered new possibilities for evaluation of breeding values for LPL. Additional to the advantages of survival analysis identified by Egger-Danner (1993), the programs also offer possibilities to adjust for time-dependent effects and to adapt the model easily without changing any source code. In this study both additional possibilities were used. Additional modelling is easier because the model can be defined in a parameter file.

East-German data collected before 1991 and daughters of East-German non-Holstein sires were excluded from the evaluation. The reason for this is that the East-German population was upgrading quickly due to the EU quota system and the introduction Holstein genes to the dairy cow population on a large scale. Vollema and Groen (1996) found systematic differences in heritabilities of longevity traits in the population of Black-and-White cows in the Netherlands before, during and after a similar process of upgrading, indicating that changes in the population structure due to the implementation of the EU quota system and the introduction of Holstein genes on a large scale affected genetic parameters. Results of genetic evaluation with complete data from East-Germany could be even more unstable, because the upgrading of the population in East-Germany was even more rapid than in the Netherlands. However, breeding value estimation of East-German sires could be carreid out by combining only a part of the East-German data with the data set from West-Germany. The regional differences in average sire EBVs after the initial run are not due to inclusion of East-German data in the evaluation. The average sire EBVs of the East-German regions did not differ substantially from the average over the whole population.

Although breeding values were estimated over all breeds, breed of the cow was included in the test-runs. The reasons for this are that breed effects on functional longevity cannot be ignored and that the proportion of Red-Holstein cows is regionally different. An implicit assumption, made by inclusion of region in the model, is that regional differences in culling are caused by environmental effects only. This is not the case because breed compositions are regionally different. By including breed in the model, the assumption is limited to regional differences in culling within breeds being caused by environmental effects only. Adjusting the sire solutions back for breed afterwards according to the amount of daughter information available per breed causes, however, that estimated breeding values are comparable across breeds.

The results of the first test-run showed different estimates of lactation number * stage, and relative yield * year-season across regions. Estimates suggested a substantial interaction of region with relative yield * year-season, and a slight interaction of region with lactation number * stage. Therefore, the interaction between region and relative yield * year-season was kept in the model for the second test-run.

The genetic trend in LPL is stable up to 1985. Bulls born from 1988, which generally still have many daughters with censored records, have a higher average RZN. This may be due to an imperfect projection of censored records, when evaluations in other dairy cow populations show a similar trend for younger sires. The low averages of RZN for birth years 1986 and 1987 may be due to daughters of these sires relatively often being compared with daughters of the younger sires. evaluations in other dairy cow populations show a similar trend for younger bulls

The influence of including region in the model in interaction with relative yield * year-season is considerable. The latter is not surprising when sires are used regionally, because the extent of culling for yield is also dependent on whether the amount of roughage available is a limiting factor. This may be regionally different as intensity of milk production is regionally different. Therefore, relative yield * year-season should be considered in interaction with region.

Acknowledgement

The authors are indebted to Vincent Ducrocq for providing the SURVIVAL-KIT V3.0 and for valuable help to get the system running.

References

Ducrocq, V. and Casella, G. (1996): A Bayesian analysis of mixed survival models. *Genet. Sel. Evol*, 28 : 505-529.

Ducrocq, V. and Sölkner, J. (1994): The Survival Kit - a Fortran package for the analysis of survival data. *Proc. 5th World Congress on Genetics Applied to Livestock Production* 22: 51-52.

Ducrocq, V. and Sölkner, J. (1998a): The Survival Kit - a Fortran package for the analysis of survival data. *Proc. 6th World Congress on Genetics Applied to Livestock Production* 27: 447-448.

Ducrocq, V. and Sölkner, J. (1998b): Implementation of a routine breeding value evaluation of dairy cows using survival analysis techniques. *Proc. 6th World Congress on Genetics Applied to Livestock Production* 23: 359-362.

Egger-Danner, C. (1993): *Zuchtwertschätzung für Merkmale der Langlebigkeit beim Rind mit Methoden der Lebensdaueranalyse.* PhD-Thesis Univ. Agric., Vienna, Austria.

Vollema, A.R. and Groen, A.F. (1996). Genetic parameters of longevity traits of an upgrading dairy cattle population. *J. Dairy Sci.*, 79: 2261-2267.