

Accuracy of Methods for the Genetic Evaluation of Survival in Dairy Cattle

J. Jamrozik, J. Fatehi and L.R. Schaeffer

*Centre for Genetic Improvement of Livestock,
Department of Animal and Poultry Science
University of Guelph, Guelph, ON, N1G 2W1, Canada*

Introduction

Many different terms and statistical models have been used to describe the likelihood that a cow will remain in the herd longer than her contemporaries. The proportional hazard (PH) model (Ducrocq *et al.*, 1998) has been adopted by a majority of countries. A second model is a multiple-trait (MT) animal model, with traits defined as a survival (0 or 1) to certain end-points or within certain time intervals in a cow's lifetime. A longitudinal generalization of multiple-trait models for survival is the random regression (RR) model proposed by Veerkamp *et al.* (1999).

Objectives of this research were to compare MT, RR and PH models for genetic evaluation of survival in dairy cattle through a simulation study. Models were ranked according to the correlation between true breeding values (TBV) and estimated breeding value (EBV) for survival traits. Predictive abilities of models were assessed through correlations between EBV and proportion of sire's daughters (PSD) that survived to a given time.

Material and Methods

Simulation:

The empirical survival curve for the Canadian Jersey breed in Figure 1 describes the probability that a cow is alive in a given month after the first calving. The same Figure shows the plot of genetic variance for survival of Jersey cows in months 1 – 100 after the first calving, estimated from the random regression model (Galbraith, 2003).

Those two curves were the input parameters for the simulation of survival data.

Permanent environmental (PE) effect for survival was assumed to have the same variance as for genetic effects. Heritability for survival was constant in each month after the first calving. Average probability that an animal was culled in the *i*-th month (average empirical hazard) was calculated as $H(i) = 1 - S(i)$, where $S(i)$ was the value of survival function for the *i*-th month.

Two phenotypes were considered, and one record for each trait was generated for every animal. Trait X (approximating a continuous production trait) was normally distributed with mean 0, heritability of 0.3 and phenotypic variance equal to 100. The second, uncorrelated trait, Y, was the time of culling of an animal, defined in the interval from 1 to 100 months after the first calving. Simulation of phenotypes for X consisted of generating additive genetic and residual components from standard normal distributions. For time of disposal, Y, overall additive genetic (*a*), PE (*p*) and residual (*e*) components for hazard were simulated for each animal as variables from normal distributions with mean zero and variance equal to 1. An animal-specific threshold for the hazard in the *i*-th month was defined as: $T(i) = H(i) + a \cdot \sigma_g(i) + p \cdot \sigma_{PE}(i) + e \cdot \sigma_e(i)$, where $\sigma_g(i)$, $\sigma_{PE}(i)$ and $\sigma_e(i)$ were genetic, PE and residual standard deviations for hazard in the *i*-th month, respectively. Variable $G(i) = a \cdot \sigma_g(i)$ will be referred to as TBV for hazard (survival). A random deviate (*d*) from standard normal distribution $N(0,1)$ was subsequently generated for a given animal. The month of culling for an animal was defined as the smallest *i* for which the value of $T(i)$ was not smaller than *d*.

Three levels of heritability (0.100, 0.050 and 0.025) of survival and two levels of number of females per generation (2000 and 4000) were considered in the simulation.

Number of males per generation was equal to 100 in each case. Nineteen generations of selection and mating were made after the base population. The length of each selection and mating cycle was set to 13.2 months, which was the length of calving interval for Canadian Jerseys (Miglior and Van Doormaal, 2005). Selection of males and females was based on phenotypes for trait X. Survival was therefore independent of X and can be treated as a functional survival. Sires were assigned to dams randomly. The best sires were selected across generations regardless of age. Cows were used as dams when they were still alive and were the top females in the population for trait X. Each mating resulted in one progeny with sex ratio equal to 0.5. Twenty-one replicates were generated for each of six scenarios.

Estimation of breeding values:

The models were:

PH model: $h_{jkl}(t) = h_0(t) \exp(\mu_j + s_l)$, where $h_{jkl}(t)$ was the hazard of k-th daughter of l-th sire, $h_0(t)$ was a Weibull baseline hazard function, μ_j was fixed effect of j-th generation, s_l was random (normally distributed) effect of l-th sire. Data for this model were month of culling for cows. Censored records were included.

MT model: $y_{ijk} = \mu_{ij} + a_{ik} + e_{ijk}$, where y_{ijk} was observation for survival (0 or 1) of the i-th trait of k-th cow, μ_{ij} was a fixed effect of j-th generation for i-th trait, a_{ik} was random additive genetic effect of k-th cow for i-th trait, e_{ijk} was a random residual effect. All random effects of the model were normally distributed. Traits ($i = 1$ to 5) were: survival from first calving to 120 DIM, survival from 120 to 240 DIM, survival from 240 DIM to second calving, survival through second lactation and survival through third lactation, following the Canadian model for survival (Sewalem *et al.*, 2007). Survival in each of the 5 traits was coded as 0 if the cow was culled during that time period, as 1 if the cow survived during that time period, and as missing if the cow did not have an opportunity to begin or complete given time period (censored data). Observations for different traits were correlated through genetic and residual effects.

RR model:

$$y_{jkt} = \sum_{m=0}^4 \beta_{jm} z_{tm} + \sum_{m=0}^4 \alpha_{km} z_{tm} + \sum_{m=0}^4 \rho_{km} z_{tm} + e_{jkt}$$

where y_{jkt} was survival observation (0 or 1) on the k-th cow at month t after first calving; β_{jm} were fixed regression coefficients specific to j-th generation; α_{km} were random additive genetic coefficients specific to k-th cow; ρ_{km} were random PE coefficients specific for k-th cow; e_{jkt} was the random residual effect for each observation, and z_{tm} were covariates. All random effects of the model were normally distributed. Orthogonal Legendre polynomials of order 4 were used for all fixed and random regressions. Each cow had up to 100 binary observations for each of 100 months after first calving. Censored records were treated as missing data in the model. Residuals for different months were assumed uncorrelated and different residual variance was allowed for each month after first calving.

Parameters for each model were estimated using the same models and the first replicate of the simulated data, for three levels of heritability and the scenarios with 2000 females per generation only. Estimates were kept constant for genetic evaluation models (replicates 2 to 21). Survival Kit (Ducrocq and Sölkner, 1994) was used for the PH model. Parameters for MT and RR models were estimated using Gibbs sampling as posterior means of 100,000 samples after 20,000 burn-in iterations. Standard Gaussian assumptions were applied to Bayesian specification of MT and RR models. Estimates of breeding values for sires for all scenarios and replicates were obtained for all three models.

Comparison of models:

Two groups of sires, with minimums of 20 or 100 daughters with survival data were selected for model comparison purposes. Correlations between TBV and EBV were used as a measure of accuracy of genetic evaluation, averaged across replicates. EBV were calculated for five different time points in cow's life, corresponding to thresholds in the MT model. They were: 120 DIM after first calving, 240 DIM after first calving, 2nd, 3rd

and 4th calvings. TBV for sire q and time t was equal to $a_q * \sigma_g(t)$. EBV were solutions for:

$$s_q, a_{tq} \text{ and } \sum_{m=0}^4 \alpha_{qm} z_{tm}, \text{ for PH, MT and RR}$$

models, respectively. TBV for survival corresponded to probability of culling at a given time. EBV from MT and RR models were probabilities of survival of a cow at a given point in time. EBV from the PH model represented the probability of culling for a cow.

Another comparison criterion was based on PSD that survived to each of the five described end-points. Those correlations can serve as a tool to examine predictive abilities of models. PSD were correlated with EBV for survival within model and simulation scenario for two groups of sires (minimum of 20 or 100 daughters).

Results and Discussion

Results in Table 1 are for heritability of 0.050 and number of females per generation equal to 2000. The highest correlations (0.53 and 0.68) were for the PH model, followed by RR and MT evaluations. MT and RR models gave practically the same accuracy for small progeny groups. The values of PH correlations were the same for all five time points while correlations for the other two models increased (in absolute terms) with time. Rankings of models were similar for all heritability levels. Accuracy of genetic evaluation for survival decreased with decreasing heritability. Larger progeny groups gave better accuracy of sire evaluation. Accuracy of sire evaluation increased with time for RR and MT models. Evaluations for survival to 120 DIM were the least accurate while evaluations for survival to 4th calving showed the highest level of accuracy in the case of MT and RR models. Accuracy of the RR model reached the level of the PH model for the survival to 4th lactation. Larger female population size resulted, in general, in slightly better accuracy of genetic evaluation for all five survival traits.

Estimates of correlation between EBV from RR, MT and PH models and PSD that survived to each of the five end-points are shown in Table 2. Rankings of models were different than those obtained from correlations between TBV and EBV. The RR model was better, with the exception of 120 and 240 DIM and 100 daughters per sire, when the highest correlations were from the MT model. Model MT gave better prediction of survived daughters than the PH model for the first three time points in the cow's lifetime. Correlations from the PH model showed relatively large changes in time compared with less variable correlations for the two other models. Similar rankings of models were obtained for the two other heritability levels. Correlations between PSD and EBV decreased with decreasing heritability level. Correlations were larger for sires with larger number of daughters. There were no differences between prediction ability of models for scenarios with 2000 or 4000 females per generation.

Conclusions

Stochastic simulation indicated that the PH model with the Weibull hazard function was the most accurate in terms of correlation between TBV and EBV for analysis of survival data in dairy cattle. RR and MT models gave lower accuracies of genetic evaluation, especially for survival in early stages of cow's lifetime. Predictive ability of models, measured by correlation between EBV and the PSD that survived to certain age after first calving, favoured the RR model.

Acknowledgements

The authors are grateful to DairyGen and the Natural Science and Engineering Research Council for their financial support.

References

- Ducrocq, V., Quaas, R.L., Pollack, E.J. & Casella, G. 1988. Length of productive life of dairy cows. 1. Justification of a Weibull model. *J. Dairy Sci.* 71, 3061-3070.
- Ducrocq, V. & Sölkner, J. 1994. "The survival kit", a FORTRAN package for the analysis of survival data. *Proc. 5th WCGALP*, 22, 51-52.
- Galbraith, F. 2003. Genetic evaluation of longevity in Ayrshire and Jersey dairy cattle using a random regression model. *M. Sc. Thesis*, University of Guelph.
- Miglior, F. & Van Doormaal, B.J. 2005. Phenotypic trends of reproductive traits in Canadian dairy breeds. *A Report to the Dairy Cattle Breeding and Genetics Committee and the Genetic Evaluation Board*, September 2005.
- Sewalem, A., Miglior, F., Kistemaker, G.J., Sullivan, P., Huapaya, G. & Van Doormaal, B.J. 2007. Modification of genetic evaluation of herd life from three-trait to a five-trait model in Canadian dairy cattle. *J. Dairy Sci.* 90, 2025-2028.
- Veerkamp, R.F., Brotherstone, S. & Meuwissen, T.H.E. 1999. Survival analysis using random regression models. *Interbull Bulletin* 26, 36-40.

Figure 1. Empirical survival curve and genetic variance of survival, by month after first calving.

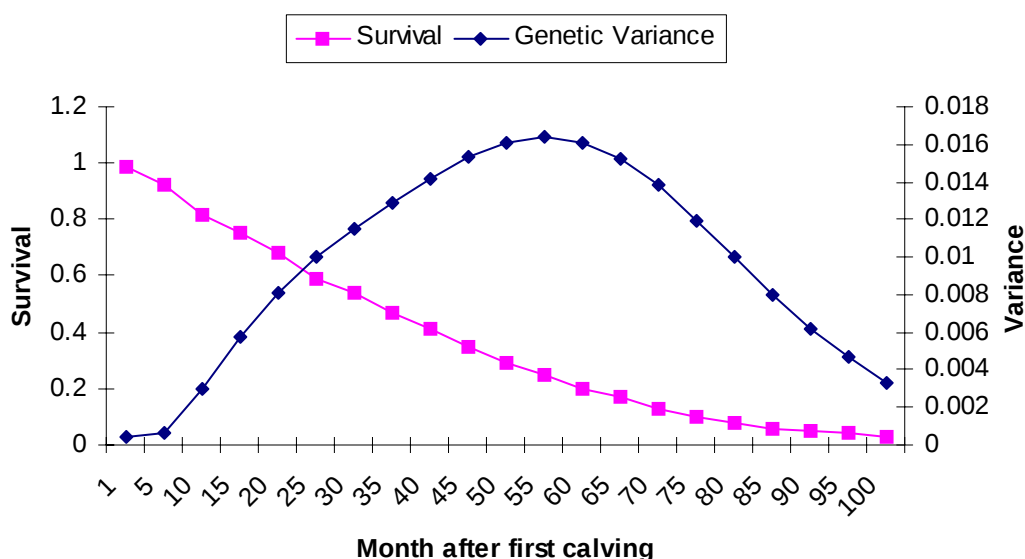


Table 1. Correlation coefficients (SD in brackets) between EBV from random regression (RR), multiple-trait (MT) and proportional hazard (PH) models and TBV for five end-points, $h^2 = 0.050$, two levels of minimum number of daughters per sire (MIN=20 and MIN=100) and number of females per generation equal to 2000.

End-point	Model					
	MT		RR		PH	
	MIN=20	MIN=100	MIN=20	MIN=100	MIN=20	MIN=100
120 DIM	-0.36 (0.09)	-0.44 (0.21)	-0.35 (0.06)	-0.50 (0.19)	0.53 (0.05)	0.68 (0.11)
240 DIM	-0.46 (0.07)	-0.58 (0.14)	-0.41 (0.06)	-0.57 (0.16)		
2nd calving	-0.46 (0.08)	-0.57 (0.15)	-0.44 (0.06)	-0.60 (0.14)		
3rd calving	-0.48 (0.07)	-0.58 (0.15)	-0.47 (0.05)	-0.65 (0.12)		
4th calving	-0.45 (0.06)	-0.57 (0.09)	-0.48 (0.05)	-0.66 (0.10)		

Table 2. Correlation coefficients (SD in brackets) between EBV from random regression (RR), multiple-trait (MT) and proportional hazard (PH) models and percentage of sire daughters that survived to each of the five end-points, $h^2 = 0.050$, two levels of minimum number of daughters per sire (MIN=20 and MIN=100) and number of females per generation equal to 2000.

End-point	Model					
	MT		RR		PH	
	MIN=20	MIN=100	MIN=20	MIN=100	MIN=20	MIN=100
120 DIM	0.66 (0.03)	0.83 (0.05)	0.73 (0.04)	0.77 (0.06)	-0.28 (0.07)	-0.40 (0.14)
240 DIM	0.64 (0.03)	0.81 (0.06)	0.68 (0.04)	0.77 (0.08)	-0.44 (0.04)	-0.56 (0.14)
2nd calving	0.66 (0.02)	0.83 (0.07)	0.76 (0.02)	0.85 (0.04)	-0.56 (0.03)	-0.71 (0.06)
3rd calving	0.68 (0.03)	0.84 (0.05)	0.84 (0.02)	0.92 (0.03)	-0.70 (0.03)	-0.83 (0.04)
4th calving	0.63 (0.04)	0.77 (0.07)	0.80 (0.02)	0.89 (0.04)	-0.73 (0.02)	-0.85 (0.03)