

Validation of genetic evaluation for longevity with a random regression model

M.L van Pelt¹, G. de Jong¹ & R.F. Veerkamp²

¹ *CRV, P.O. Box 454, 6800 AL Arnhem, the Netherlands
mathijs.van.pelt@crv4all.com (Corresponding Author)*

² *Wageningen University & Research Animal Breeding and Genomics, P.O. Box 338, 6700 AH, Wageningen, the Netherlands*

Summary

A new model for the genetic evaluation for longevity was developed in the Netherlands based on a random regression animal model (RRM). The existing system for genetic evaluation was based on a proportional hazard model (PHM). Changes adopted with the RRM were 1) fitting multiple genetic effects across the life of a cow, 2) animal model vs. sire-mgs model, 3) adjusting for milk production at herd level, and 4) fitting fixed effects differently. The aim here was to evaluate and validate the new RRM and compare the EBV with the current PHM. For the new and existing model 11 evaluation runs were performed with one year of data from 2007 up to 2017 added in every new run. EBV for survival until 72 months were calculated for all animals. Stability of breeding values was analysed as the difference with EBV-2017 and the correlation of the first EBV of a bull with a later EBV. The trait analysed was survival per month, fitted with a fifth-order Legendre polynomial until 72 months after first calving. EBV were overestimated mainly in EBV-runs based on the earlier years, due to incomplete daughter information. Adjusting for within-herd production level reduced this bias. Based on the correlation between first and later EBV, the ranking of bulls was shown to be more stable for RRM than for PHM. RRM with adjustment for milk yield is the preferred model for longevity, as it resulted in more stable ranking of bulls with smallest overestimation of EBV based on incomplete daughter information.

Keywords: longevity, validation, genetic evaluation, random regression

Introduction

Longevity is a complex trait; true longevity is only available at the end of a cow's life, whereas selection decisions are made early in life. The proportional hazard model (PHM) can handle censored data, fit time-dependent fixed effects and account for the non-normal distribution of the survival data. However, practical experience in the Netherlands and Flanders has shown that EBV fluctuated more than expected from the change in reliability. Veerkamp et al. (2001) proposed the random regression model (RRM); similar to PHM, RRM can handle censored data and use time-dependent fixed effects. Moreover, RRM can fit multiple genetic effects and it is computationally feasible to fit an animal model on large data sets. Non-unity genetic correlations for survival in different months and heterogeneous genetic variances can be modelled over the life of a cow (Van Pelt et al., 2015). Age at first calving (AFC) and within-herd production level need to be fitted time-dependent (Van Pelt et al., 2016a). Survival adjusted for within-herd production level resulted in a trait that is genetically more constant over time (Van Pelt et al., 2016b). The objective of this study was to develop a new genetic

evaluation for longevity based on a RRM, and validate whether EBV of this RRM are more stable compared to EBV from the current PHM.

Material and methods

Data

Length of productive life was defined as the time from first calving to the last test date for milk production, before the animal died or was culled for slaughter; this also included dry periods. For PHM, total length of productive life was analysed; for RRM, the analysed period was survival until 72 mo after first calving. Data were available from the Dutch/Flemish cattle improvement cooperative CRV (CRV, Arnhem, the Netherlands). The data set was constructed from records of pedigree, lactations and movements of cows in the Netherlands. Herdbook-registered cows from a dairy breed with a test-day record on or after January 1, 1988 were included. Data up to February 15, 2017 were used. Cows were required to be at least 640 d old at first calving. If the first calving of a cow took place before the starting date of the study, the record was considered to be left-truncated. Records of cows that were still alive at the time of data collection were considered to be right-censored. Records of cows that were moved to another milking herd were also considered to be right-censored, if this herd was not participating in a milk recording scheme.

Records for RRM were constructed for each month a cow was present in a herd, from first calving up to the month the cow was culled, or 72 mo, or when the cow was censored. A cow culled in month j has $j - 1$ records with score 100 (alive), and record j with score 0 (culled). Monthly records were treated as missing after culling. Additionally, with RRM, cows with an AFC of >40 mo were deleted. The total data set for RRM comprised 370,871,367 records from 10,924,641 animals in 48,407 herds. The data set for PHM comprised 10,939,556 animals. For both RRM and RRM a voluntary waiting period of 270 days applied; the information was included at a time when the cow could have been productive for at least 270 days after the first calving.

Statistical Model

Changes in the new RRM compared to the current PHM were 1) fitting multiple genetic effects across the life of a cow, 2) animal model vs. sire-mgs model, 3) adjusting for milk production on herd level, and 4) fitting fixed effects differently. The RRM was a linear random regression animal model and survival per month was analysed. Fixed effects included herd-year-season-lactation-stage with year-season of first calving, year-season-AFC-production-lactation-stage with production as within-herd production level, and herd size change; covariables were fitted for heterosis and recombination; the additive genetic effect for the animal was fitted as a random effect with a fifth order Legendre polynomial. Two models were used for RRM, with and without within-herd production level (RR-func and RR-true). Within-herd production level was fitted to correct for culling due to low production, which is assumed to be the major source of voluntary culling yielding EBV for functional longevity, whereas not including within-herd production level yielded EBV for true longevity similar to the EBV from the PHM. For PHM, a piecewise Weibull PHM was used in the current genetic evaluation, as described in Van der Linde et al. (2004, 2007). The RRM yielded EBV per month; an overall EBV over 72 mo was constructed by calculating the area under the survival curve for an animal and the population mean. The difference between these areas yields the overall EBV. Overall heritabilities for

longevity were 0.12 for RR-func, 0.15 for RR-true, and 0.12 for PHM. Genetic standard deviations were 7.1 mo for RR-func, 7.6 mo for RR-true, and 9.0 mo for PHM.

Validation

In total, three models were examined: RR-func, RR-true, and PHM. To evaluate the stability of breeding values for bulls, subsets of data for all three models were used in the genetic evaluations by deleting most recent years. The national data set contained data up to February 15, 2017. First a genetic evaluation was performed using all data (full run, EBV-2017). Following this, 10 historic runs were performed by deleting one year of data for each subsequent run. The first historic run involved data from January 1, 1988 to February 15, 2016 (EBV-2016); the tenth run involved data to February 15, 2007, (EBV-2007). Per model, a bull could have a maximum of 11 breeding values for longevity, one from each evaluation (i.e., the full run and 10 historic runs), depending on when a bull received the first proof based on progeny.

For black and white Holstein bulls born since 2001 and tested in the Netherlands and Flanders, the first EBV for both PHM and RRM was defined in the evaluation run where ≥ 10 daughters were at 10 mo after first calving and ≤ 50 daughters were ≤ 24 mo after first calving. Test bulls and proven bulls were validated separately. Test bulls were defined as bulls with ≤ 250 daughters in any of the evaluation runs, and proven bulls as those with > 250 daughters for EBV-2017. Mean difference with EBV-2017 and the correlation between first and later EBV of sequential runs were calculated. Genetic trends were calculated based on all black and white Holstein bulls, i.e. not only nationally tested bulls. These runs were not sequential in terms of yearly evaluations, but represented the sequential run after the first evaluation of a bull.

Results and discussion

A total of 2,378 (263) test (proven) bulls had at least one EBV-run, and 290 (21) test (proven) bulls had eleven EBV-runs. The first EBV of test bulls was on average overestimated by 0.15 genetic standard deviations (gsd) for RR-func, 0.28 gsd for RR-true, and 0.26 gsd for PHM (Figure 1). For all models the overestimation reduced when more information was added in later EBV-runs. For proven bulls, the first EBV a bull receives was on average overestimated by 0.20 gsd for RR-func, 0.30 gsd for RR-true, and 0.20 gsd for PHM (Figure 2). Overestimation of EBV reduced fastest for RR-func for both test bulls and proven bulls, below the level of 0.05 gsd in for the third EBV a bull receives. Adjusting for within-herd production level reduced the overestimation, when RR-func is compared with RR-true. For both test bulls and proven bulls, RR-func resulted in the smallest overestimation of first EBV. The overestimation that was remaining, is likely due to incomplete daughter information, i.e. not all daughters were able to pass 72 mo after first calving.

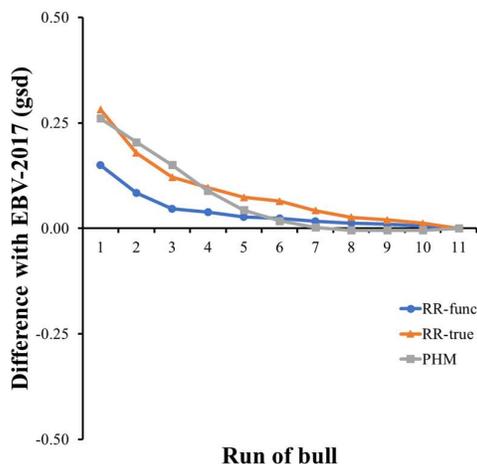


Figure 1. Mean difference of n^{th} EBV with EBV-2017 for test bulls.

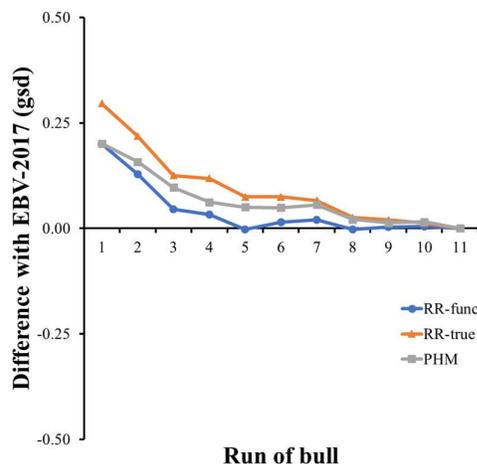


Figure 2. Mean difference of n^{th} EBV with EBV-2017 for proven bulls.

Correlations between first and later EBV show that for both test bulls and proven bulls the correlations of RR-func were highest and were lowest for PHM (Table 1). For RR-func, after 4 EBV-runs the correlation with first EBV was almost stable (~ 0.70 for test bulls), whereas the correlation kept declining for PHM (~ 0.65) and RR-true (~ 0.60). For all three models, reranking occurred going from first to later EBV. Ranking of bulls was most stable for RR-func, and least stable for PHM.

Table 1. Correlations between first and later EBV for RR-func, RR-true and PHM for test bulls (<250 daughters) and proven bulls (≥ 250 daughters) born between 2001 and 2011.

n^{th} EBV	Test bulls			Proven bulls		
	RR-func	RR-true	PHM	RR-func	RR-true	PHM
2	0.86	0.85	0.83	0.86	0.84	0.79
3	0.77	0.76	0.75	0.75	0.70	0.68
4	0.73	0.71	0.68	0.72	0.67	0.65
5	0.72	0.69	0.65	0.63	0.62	0.58
6	0.71	0.68	0.63	0.60	0.63	0.56
7	0.70	0.68	0.61	0.61	0.63	0.59
8	0.69	0.65	0.59	0.62	0.59	0.58
9	0.69	0.65	0.58	0.58	0.57	0.52
10	0.68	0.65	0.59	0.50	0.56	0.46
11	0.69	0.68	0.63	0.75	0.72	0.60

Genetic trends for all black and white Holstein bulls are shown for EBV-2008, EBV-2012 and EBV-2016 for all three models (Figure 3-5). When no bias is present in the genetic evaluation genetic trends should be overlapping for the different EBV-runs. None of the three models showed completely overlapping genetic trends from the three different EBV-runs. However, visual inspection showed that for RR-func and PHM until birth year 2000 the differences between EBV-runs were small, followed by diverging genetic trends. For RR-true, mean EBV in older birth years were still changing and in more recent birth years changes in mean EBV were bigger compared to the other two models. For birth year 2003, the overestimation in mean EBV (EBV-2016 – EBV-2008) was 0.20 gsd for RR-func, and 0.46 gsd for RR-true, and 0.57 gsd for PHM. For birth year 2000, the EBV were underestimated by 0.05

gsd for RR-func, and the overestimation of EBV reduced 0.13 gsd for RR-true, and 0.16 gsd for PHM. The overestimation in genetic trends were in line the Figure 1 and 2, and also showed that overestimation was reduced almost completely within 3 years for RR-func, where PHM needed 5 years, and for RR-true the overestimation remains.

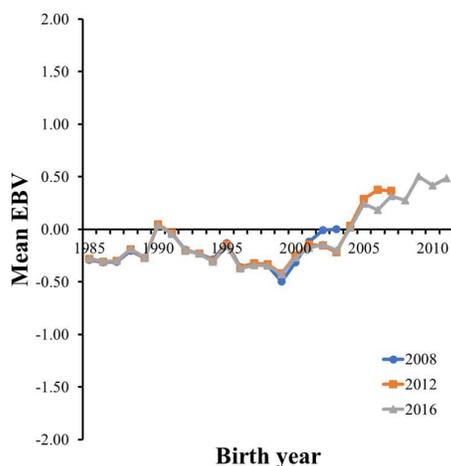


Figure 3. Estimates of genetic trends for longevity from RR-func for Holstein bulls using data until 2008, 2012 and 2016.

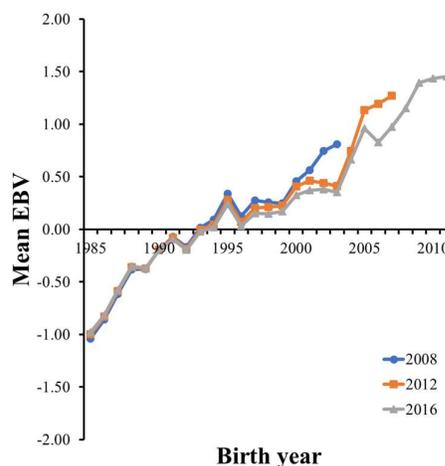


Figure 4. Estimates of genetic trends for longevity from RR-true for Holstein bulls using data until 2008, 2012 and 2016.

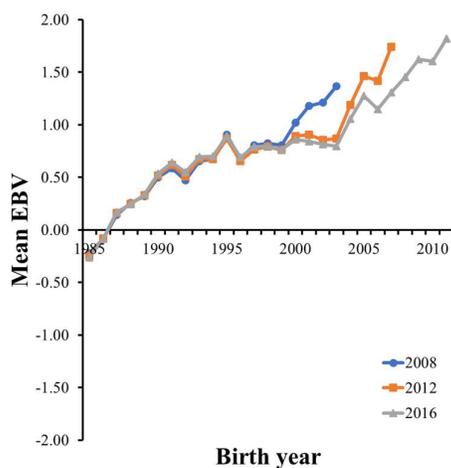


Figure 5. Estimates of genetic trends for longevity from PHM for Holstein bulls using data until 2008, 2012 and 2016.

Conclusion

A new genetic evaluation model for longevity was developed based on a random regression animal model. EBV were overestimated mainly in first EBV-runs on the earlier years for RR-true and PHM, due to incomplete daughter information. Adjusting for within-herd production level reduced this bias for RR-func. Based on the correlation between first and later EBV, the ranking of bulls was shown to be more stable for RRM than for PHM. RRM with adjustment

for milk yield is the preferred model for longevity, as it resulted in more stable ranking of bulls with smallest overestimation of EBV based on incomplete daughter information.

List of References

- Van der Linde, C., G. de Jong, and A. Harbers. 2004. Using a piecewise Weibull mixed model in the genetic evaluation for longevity. *Interbull Bulletin* 32:157.
- Van der Linde, C., A. Harbers, and G. de Jong. 2007. From functional to productive longevity in the Netherlands. *Interbull Bulletin* 37:203.
- Van Pelt, M. L., G. de Jong, and R. F. Veerkamp. 2016a. Changes in the genetic level and the effects of age at first calving and milk production on survival during the first lactation over the last 25 years. *Animal* 10:2043-2050.
- Van Pelt, M. L., V. Ducrocq, G. de Jong, M. P. L. Calus, and R. F. Veerkamp. 2016b. Genetic changes of survival traits over the past 25 years in Dutch dairy cattle. *J. Dairy Sci.* 99:9810-9819.
- Van Pelt, M. L., T. H. E. Meuwissen, G. de Jong, and R. F. Veerkamp. 2015. Genetic analysis of longevity in Dutch dairy cattle using random regression. *J. Dairy Sci.* 98:4117-4130.
- Veerkamp, R. F., S. Brotherstone, B. Engel, and T. H. E. Meuwissen. 2001. Analysis of censored survival data using random regression models. *Anim. Sci.* 72:1-10.