Improving the Genetic Evaluation for Longevity in the Netherlands

M.L. van Pelt[1], G. de Jong[1] and R.F. Veerkamp[2]
[1]*CRV, P.O. Box 454, 6800 AL Arnhem, the Netherlands*
[2]*Wageningen University & Research Animal Breeding and Genomics, P.O. Box 338, 6700 AH, Wageningen, the Netherlands*
E-mail: Mathijs.van.pelt@crv4all.com

Abstract

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 from 2007 up to 2017, where one year of data was added in every new run. Stability of breeding values was analysed as the difference with EBV-2017 and the correlation of the first EBV of a bull with later EBV. The trait analysed was survival per month, fitted with a fifth-order Legendre polynomial until 72 months after first calving. EBV equated to 72 months were calculated for all animals. EBV were overestimated mainly in first EBV-runs, 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.

1. 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. Therefore it is necessary that EBV are accurately estimated and will remain stable. 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 computational 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.

2. 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 length of productive life 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. Additional procedure for RRM was that 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:

$$Y_{ijklmno} = HYS\_LS_i + YSAM\_LS_j + HSC_k + het_l$$
$$+ rec_m + \sum_{q=0}^{5} animal_{nq}\, leg_{oq}$$
$$+ rest_{ijklmno}$$

where

$Y_{ijklmno}$: observation for survival in month o after first calving; mo $1 − 72$;

HYS_LS_i : fixed effect for herd-year-season x lactation-stage i; year-season observation, lactation split in 1, 2, 3+, stage of lactation split in mo 1-2, 3-9, 10+ and dry period;

$YSAM\_LS_j$: fixed effect for year-season x AFC x within-herd production level x lactation-stage j; year-season of observation, AFC in months 21, 22, …, 34, 35+, within-herd production level is defined per 3 years and is divided in 5 classes of 20% each for predicted or realised age-corrected 305-day yield of kg fat and protein;

HSC_k : fixed effect for herd size change k; HSC is calculated by comparing the number of cows present in a herd in a year with the number of cows in the same herd one year later. Seven classes are distinguished: shrinkage between 90 and 50%, shrinkage between 50 and 30%, shrinkage between 30 and 10%, neither shrinkage nor growth over 10%, growth between 10 and 30%, growth over 30%, and herds that were terminated (more than 90% shrinkage).

het_l : covariable for heterosis l of animal n;

rec_m : covariable for recombination m of animal n;

leg_{oq} : covariates of order q Legendre polynomial for month o;

$animal_{nq}$: additive genetic random regression coefficient of animal n corresponding to polynomial q;

$rest_{ijklmno}$: random residual effect of $Y_{ijklmno}$.

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, and subtracting both areas yields the overall EBV.

Overall heritabilities for longevity were 0.12 for RRM, and 0.12 for PHM. Genetic standard deviations were 7.1 mo for RRM, and 9.0 mo for PHM.

Validation

To evaluate the stability of breeding values for bulls, subsets of data for both 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.
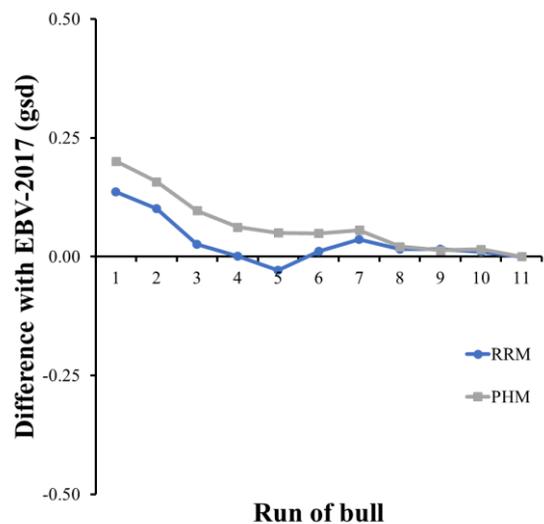
3. Results and Discussion

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 0.21

genetic standard deviations (gsd) for RRM, and 0.26 gsd for PHM (Figure 1).
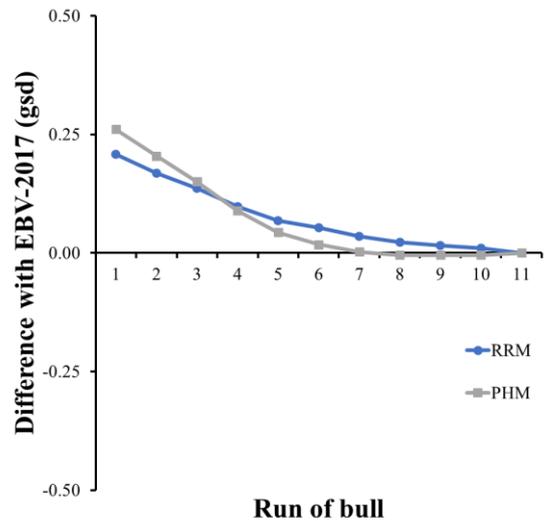
Figure 1. Mean difference of nth EBV with EBV-2017 for test bulls.

Figure 2. Mean difference of nth EBV with EBV-2017 for proven bulls.

For all models the overestimation reduced when more information was added in later EBV-runs. For proven bulls, first EBV was on average overestimated by 0.14 gsd for RRM, and 0.20 gsd for PHM (Figure 2). Proven bulls had smaller overestimation than test bulls, and this overestimation reduced faster below the level of 0.05 gsd, for RRM in the third EBV-run and for PHM in the eigth EBV-run. For both test bulls and proven bulls, RRM resulted in a smaller overestimation of first EBV compared

to PHM. 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.

Correlations between first and later EBV show that for both test bulls and proven bulls the correlations of RRM were higher compared to PHM (Table 1). For RRM, after 4 EBV-runs the correlation with first EBV was almost stable (~0.70 for test bulls), whereas the correlations were lower and kept declining for PHM (~0.65). For both models, reranking occurred going from first to later EBV. However, ranking of bulls was more stable for RRM than for PHM.

Table 1. Correlations between first and later EBV for RRM and PHM for test bulls (<250 daughters) and proven bulls (≥250 daughters) born between 2001 and 2011.

EBV-run	Test bulls		Proven bulls	
	RRM	PHM	RRM	PHM
2	0.86	0.83	0.86	0.79
3	0.77	0.75	0.75	0.68
4	0.73	0.68	0.72	0.65
5	0.72	0.65	0.65	0.58
6	0.71	0.63	0.64	0.56
7	0.71	0.61	0.64	0.59
8	0.70	0.59	0.65	0.58
9	0.70	0.58	0.62	0.52
10	0.69	0.59	0.53	0.46
11	0.71	0.63	0.75	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 and 4). When no bias is present in the genetic evaluation genetic trends should be overlapping for the different EBV-runs. Neither model showed completely overlapping genetic trends from the three different EBV-runs. However, visual inspection showed that for RRM and PHM until birth year 2000 the differences between EBV-runs were small, followed by diverging genetic trends. For birth year 2003, the overestimation in mean EBV (EBV-2016 – EBV-2008) was 0.15 gsd for RRM, and 0.56 gsd for PHM. For birth year 2000, the overestimation reduced to 0.04 for RRM, and 0.16 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 RRM, where PHM needed 5 years.
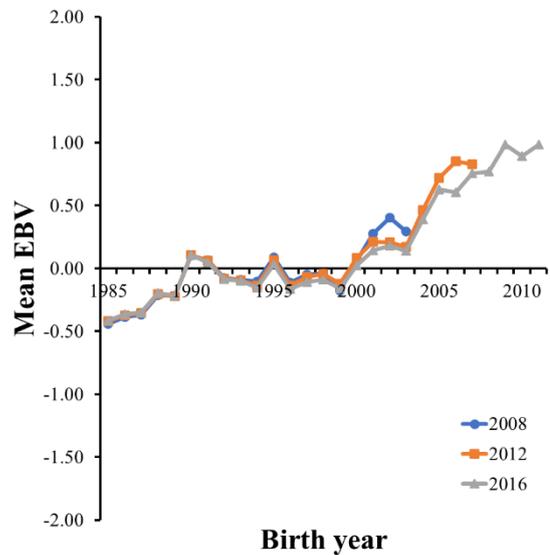
Figure 3. Estimates of genetic trends from RRM for Holstein bulls using data until 2008, 2012 and 2016.
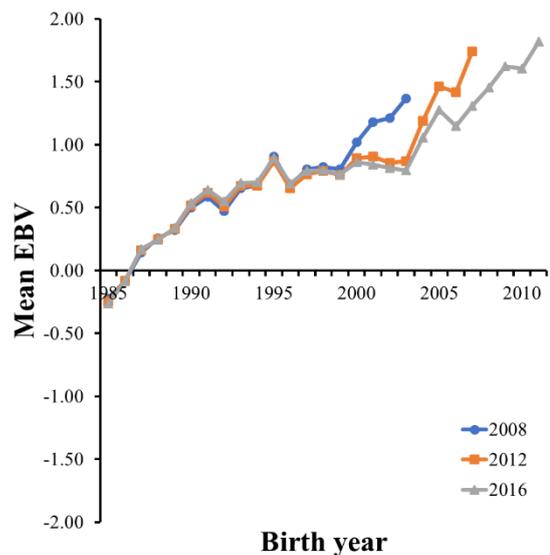
Figure 4. Estimates of genetic trends from PHM for Holstein bulls using data until 2008, 2012 and 2016.

Presentation of breeding value

The current EBV from the PHM is for productive longevity, and farmers are used to the interpretation of this EBV scaled in days. Therefore the new EBV from the RRM should be comparable with current EBV. The total EBV will be extrapolated from 72 mo to total

life. The total EBV for functional longevity over total life is converted to true longevity with an selection index including EBV for kg milk, kg fat and kg protein. Predictors traits (subclinical mastitis, claw health and locomotion) are added to increase reliability of the EBV. First results showed a correlation between EBV from RRM and PHM around 0.90.

4. 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, 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.

5. References

Van der Linde, C., de Jong, G. & Harbers, A. 2004. Using a piecewise Weibull mixed model in the genetic evaluation for longevity. *Interbull Bulletin 32,* 157-162.

Van der Linde, C., Harbers, A. & de Jong, G. 2007. From functional to productive longevity in the Netherlands. *Interbull Bulletin 37,* 203-207.

Van Pelt, M.L., de Jong, G. & Veerkamp, R.F. 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., Ducrocq, V., de Jong, G., Calus, M.P.L. & Veerkamp, R.F. 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., Meuwissen, T.H.E., de Jong, G. & Veerkamp, R.F. 2015. Genetic analysis of longevity in Dutch dairy cattle using random regression. *J. Dairy Sci. 98,* 4117-4130.

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