Genetic Analysis of Cow Survival in the Israeli Dairy Cattle Population

Petek Settar[1] and Joel I. Weller
Institute of Animal Sciences A. R. O., The Volcani Center, Bet Dagan 50250, Israel

Abstract

The linear model method of VanRaden and Klaaskate to analyze herd life was expanded. Information on conception and protein yield was included in the estimation of predicted herd life of Israeli Holsteins. Variance components were estimated by a multitrait animal model. Heritability was slightly higher for herd life than for number of lactations, but genetic correlations were close to unity. Animal model heritability estimates of herdlife were higher than previous sire model estimates. The expected herd life of pregnant cows was 420 d greater than open cows. Each kg increase in protein yield increased expected herd life by 9.5 d. Heritability of expected herd life increased from 0.11 for cows 6 mo after first calving to 0.14 for cows 3 yr from first calving. The genetic correlation of expected and actual herdlife increased from 0.87 for records cut after 6 mo to 0.99 for records cut 3 yr after first calving. Phenotypic correlations increased from 0.61 to 0.94. Sire genetic evaluations based on expected herd life of live cows were strongly biased if all records were weighted equally, while evaluations derived by weighting incomplete records to account for the effects of current herd life on variance components were nearly unbiased.

1. Introduction

Longevity, or herd life (**HL**), is of major economic importance in dairy cattle (VanRaden and Wiggans, 1995). Three basic strategies have been suggested to evaluate longevity for live cows. First, cow survival to a specific age can be analysed as a binary trait by either linear or threshold models (Boettcher, et.al., 1998, Harris,et.al., 1992, Jairath, et.al., 1998, Vollema and Groen, 1998). Second, VanRaden and Klaaskate (1993) proposed estimating life expectancy of live cows and including these records in a linear model analysis. Estimates based on incomplete data are regressed toward the mean, and therefore have lower heritability and variance than do complete records (Meijering and Gianola, 1985, VanRaden et.al., 1991, Weller, 1988). The third method is survival analysis or consideration of cows still alive as censored records (Boettcher, et.al., 1998, Emanuelson et.al., 1998, Vollema and Groen, 1998, Vukasinovic, et.al., 1997). Although the previous methods could be applied to either animal or sire models, survival analysis can only be applied to sire models, and evaluations will be biased if the number of daughters per sire with complete records is low (Vukasinovic, et.al.,1997).

There have been numerous suggestions for a definition of the longevity trait, based chiefly either on the number of parities or the actual length of HL (Vollema and Groen, 1996). Many studies have proposed analysing functional HL, which is generally computed as longevity adjusted for milk yield (Boettcher, et.al., 1998, Dekkers, 1993, Emanuelson et.al., 1998, Jairath, et.al., 1998, Strandberg and Hakansson, 1994, Vollema and Groen, 1996, Vollema and Groen, 1998). This trait accounts only for culling that is due to causes other than milk yield. The problem of double counting of production in a selection index that includes both milk yield and uncorrected longevity correlated with yield can also be handled by computing appropriate economic values for these traits (VanRaden and Wiggans, 1995). Analysing HL adjusted for yield is complicated because selection goals change over time. Until 1980, milk yield was the primary selection objective of most breeding programs. Now protein yield is the chief goal, and many countries put a negative economic weight on milk yield (Leitch, 1994).

VanRaden and Klaaskate (1993) used cumulative months in milk, current months in milk, age at first calving, current months dry for dry cows, first parity milk yield, and lactation

---

[1] Current adress: Ege University Faculty of Agriculture, • zmir,Turkey

status (dry or milking) to predict the HL of live cows.

In addition to low yield, the main causes of cow culling are mastitis and nonconception. In Israel, all milk recorded cows are checked for pregnancy 60 d after insemination unless the cow is reinseminated prior to 60 d (Weller and Ezra, 1997). Thus, the pregnancy status of all cows is known in real time, and this information can be used to increase the accuracy of HL predictions.

Objectives of this study were to measure of longevity for Israeli Holsteins; to study the effect of incorporating data on pregnancy, days open (**DO**), and protein yields in the computation of expected HL; to compute adjustment factors for live cows; and to determine the effects of the adjustment procedure for incomplete records on genetic evaluations. Adjustment factors were derived from variance components for complete and incomplete HL records estimated by a multitrait animal model analysis, with records of different lengths considered as correlated traits.

2. Material and methods

2.1 Estimation of expected HL for live cows and estimation of variance components

Preliminary data set consisting of 559,035 Israeli Holstein lactation records with first calving dates from 1984 through 1989, was generated to predict HL for live cows. Lactation records were discarded from the analysis if the first parity record was missing; age at first calving was <570 or >1000 d; any calving intervals were <250 d; mean calving interval was >500 d; last recorded parity was <7, and exit day was missing for that parity; cows with >1 parity with valid exit dates; protein yield was <50 or >600 kg for the last valid parity record; or if days in milk >500 or if a cow was scored pregnant but days open = 0. After edits, there were 51,888 valid cow records in this data set (Table 1).

A linear model was used to estimate HL from incomplete records. Multiple records were derived for each cow by cutting the records at 6-mo intervals beginning 6 mo after first calving until 4 yr after first calving and then cutting records at yearly intervals until 6 yr after first calving. At each cut date, a record was generated only for cows that survived until that cut date. Thus, up to 10 records were generated per cow. This data set included 231,458 cow records.

The dependent variable was remaining HL (**RHL**), computed as days from the cut date to the exit date. The independent variables were current HL (**CHL**), defined as days from first calving to cut date; last parity prior to cut (PPC); expected protein yield of the last parity prior to cut (**EPY**); pregnancy status (**PS**) where 1 = pregnant, and 0 = not pregnant; DO of the last lactation for cows pregnant by the cut date, or days in milk for cows not pregnant at the cut date, days pregnant (**DP**) at cut date for pregnant cows; and days dry (**DD**) at the cut date if the record was cut during the dry period. Days pregnant = 0 for cows not pregnant at the cut date. The EPY was computed as described previously (Weller, 1988). The HYS were defined relative to the first parity calving date, and were absorbed. PPC, PS, and HYS were analyzed as discrete effects; all other effects were analyzed as continuous variables.

If predicted RHL <0 then predicted RHL was set to zero. Estimated HL (**EHL**) was computed as HL for cow records that were culled prior to the cut, and as the sum of predicted RHL and CHL for records cut after the cut date. Genetic and environmental variance components among HL and EHL computed for the 10 cut dates were estimated by multitrait REML (Misztal, et.al., 1995) using the animal model for these 11 traits.

2.2. Estimation of Adjustment Factors for Genetic Evaluations of Incomplete Records

The EHL records as a function of CHL were first adjusted so that the genetic covariance between EHL and actual HL records would equal the genetic variance of the actual records. Multiplicative adjustment factors were estimated PROC NLIN of SAS (1988) based on the following nonlinear function of CHL:

$$RG_i = b0 + (b1/CHL_i) + e_I \qquad [1]$$

where RG_i is the ratio of the square root of the genetic covariance between HL and EHL_i, EHL at cut date i, b0 and b1 are regression constants, and e_i is the residual. With this formula, as CHL increases, the term b1/CHL becomes negligible with respect to b0, and b0 should be approximately equal to unity. The EHL records were then multiplied by the predicted values of RG for Equation [1], PRG, which also increased the mean. The following nonlinear model was used to adjust for the increase in the mean as a function of CHL based on data set 2:

$$EHLX_j = b2 + b3/CHL_j + e_j$$

where $EHLX_j = EHL \times PRG$, for cow j, b2 and b3 = regression constants, and the other terms are as described previously. Predicted HL (**PHL**) was

then computed as EHLX - b3/CHL; b2 was not subtracted because, as a constant, it would have the same effect on all records.

In the animal model analysis, the PHL records were weighted by the inverse of the ratio of the residual variances of PHL and HL as a function of CHL. Weighting factors for the square root of the residual variances of PHL were computed using the model of Equation [1]. If this model is appropriate, then b0 should be approximately equal to the residual standard deviation of HL.

2.3. Genetic Evaluation for Longevity

Data set 2 was generated for computation of animal model genetic evaluations for HL and PHL. The HL was computed for each cow as described previously, except that if HL >2557 d (7 yr), then the HL was set equal to 2557 d. The CHL was computed as days from first calving to January 1, 1990. Cows with CHL <35 d were deleted, leaving 45,300 cow records.

The animal model was used to compute genetic evaluations for HL, PHL with equal weights for all records, and PHL weighted as described previously. Pedigree information from all known parents and grandparents was included. The pedigree file included a total of 379 sires. Twenty-five phantom parent groups were defined by year of birth and sex of parent (Wiggans, et.al., 1988). Number of cows, records, sire, and HYS included in data set 2 are also given Table 1. The two methods for analyzing PHL were compared by correlations of the sires' EBV for these methods and the EBV for HL and by the regression of the HL sire EBV on the PHL EBV.

Genetic evaluations for HL based on the method developed were also computed on the complete Israeli-Holstein dairy cattle population in September 1998, including 284,541 cows with first calving dates since January 1, 1985, and at least 35 DIM at the evaluation date. Other edits were the same as for data set 2. The numbers of cows, records, bulls, HYS, and genetic groups for this data set (data set 3) are also given in Table 1. Genetic trends were computed as the regression of the cows' EBV on their birth dates, including all cows born since 1981. Phenotypic trends were computed as the regression of the cows' HL on their birth dates, but including only cows with valid HL records. For live cows EHL was used instead of HL to calculate the phenotypic trend. Correlations were computed between bull EBV for HL and milk and protein production for bulls with reliability > 0.5 for all three traits. Bull EBV for milk and protein were

computed by a standard animal model (Weller and Ezra, 1997).

3. Results and Discussion

The RHL was estimated from data set 1 as described previously. After removing nonsignificant effects from the analysis, the final equation for estimating RHL from the incomplete records was

$$RHL = PPC - 908 + 0.131 \times CHL + (-0.00029)$$

$$CHL^2 + (-0.872) \ DP + (-5.218) \ DO + (0.0169) \ DO^2 + (-0.0000313) \ DO^3 + (0.00126) \ CHL \times DO + (-0.536) \ DD + (9.532) \ EPY + (-0.011) \ EPY^2$$

[3]

All effects included in the final model were significant at $p < 0.0001$. The discrete parity and pregnancy status effects are given in Table 2. Parity effects show no discernible trend because this effect is highly confounded with CHL, which was also included in the model. Pregnancy at the cut date increased RHL by 420 d, but the effect DP was negative.

Genetic correlations between HL and EHL increased from 0.87 for records cut after 6 mo to 0.99 for records cut after 3 yr. The phenotypic correlation was 0.61 for records cut after 6 mo, and increased to 0.94 for records cut after 3 yr. The genetic and phenotypic correlation estimates between complete and incomplete HL records were higher than were those reported by VanRaden and Klaaskate (1993), but they did not include data on pregnancy status.

Heritability increased from 0.11 for records cut at 6 mo to 0.14 for records cut after 3 yr and then remained constant. VanRaden and Klaaskate (1993), using a sire model, reported that heritability of cut records increased from 0.03 to 0.08.

The genetic covariances between the complete records and incomplete records were then used to compute genetic adjustment factors based on Equation [1]. The coefficients are given in Table 3. The constant coefficient was equal to 656, and the square root of the residual variance of HL was 666. For convenience, the residual weighting factors were then computed as 656/(656 +45256/CHL). Although both the genetic covariances and the environmental variances were monotonic functions of CHL, the ratio of the residual variances after adjustment to equal genetic covariances is no longer monotonic. The

coefficients of determination for all three nonlinear models are also given in Table 3. The coefficient of determination was lowest for the residual variance of PHL, but all values were >0.9.

The EBV for all animals included in data set 2 were computed for HL and PHL with all records cut at January 1, 1990. Correlations between sire EBV for HL and PHL with and without weighting factors are given in Table 4 for all sires and for sires with >10 daughters. Because evaluations of young sires should be most effected by incomplete records, correlations are also given for sires born after 1981, 1982, 1983, and 1984. Correlations between HL and PHL computed with and without weighting factors were 0.94 and 0.93, respectively. Correlations were lower if only bulls with >10 daughters were included, and were again marginally higher for PHL computed with weighting factors. As expected, correlations decreased with decreases in the bulls' ages. In all cases differences between correlations for PHL computed with and without weighting factors were no more than 1%.

The regressions of sire EBV for HL on PHL with and without weighting factors are also given in Table 4. Without weighting factors, slopes for all bulls were about 1.3 but were nearly equal to unity for evaluations computed with weighting factors. Thus, nearly unbiased evaluations are derived with weighting factors, and evaluations based on equal weights of all records are biased. For the young sires, without weighting factors, regressions increased up to 1.75 for bulls born after 1984 and were, therefore, highly biased. With weighting factors, regressions decreased slightly with the bull's age but were still 0.85 for bulls born after 1984. Thus bias was much smaller with weighting factors.

In the analysis of the complete Israeli Holstein population the phenotypic trend for HL was –15 d/yr, and the genetic trend was 9 d/yr. The genetic correlations between the sire EBV for HL and milk and protein production by birth year of bulls born since 1986 are given in Table 5. The correlations for all bulls born since 1986 are also given. Correlations were lowest in 1988 and 1989, but no clear trends are evident. In our analysis, HL was not adjusted for milk production, and EPY was used to predict HL for live cows. This should tend to increase the similarity between EBV for production and HL, especially for young sires whose daughters are in first lactation. However, the genetic correlations between protein and HL were nearly equal for sires born in 1993, as compared to sires born in 1986.

References

Boettcher, P. J., L. K. Jairath, and J.C.M. Dekkers. 1998. Alternative methods for genetic evaluation of sires for survival of their daughters in the first three lactations. *Proc. 6th World Cong. Genet. Appl. Livest. Prod.*, Armidale, Australia 23:363-366.

Dekkers, J.C.M. 1993. Theoretical basis for genetic parameters of herd life and effects on response to selection. *J. Dairy Sci.* 76:1433-1443.

Emanuelson, U., J. Carvalheira, P. A. Oltenacu, and V. Ducrocq. 1998. Relationships between adjusted length of productive life and other traits for Swedish dairy cattle. *Proc. 6th World Cong. Genet. Appl. Livest. Prod.*, Armidale, Australia 23:367-370.

Harris, B. L., A. E. Freeman and E. Metzger. 1992. Analysis of herd life in Guernsey dairy cattle. *J. Dairy Sci.* 75:2008-2016.

Jairath, L., J.C.M. Dekkers, L. R. Schaeffer, Z. Liu, E. B. Burnside, and B. Kolstad. 1998. Genetic evaluation for herd life in Canada. *J. Dairy Sci.* 81:550-562.

Leitch, H. W. 1994. Comparison of international selection indices for dairy cattle breeding. *Proc. Open. Session Interbull Annu. Meet.* Ottawa, Canada, 10: (unnumbered). International Bull Evaluation Service. Department of Animal Breeding and Genetics, SLU, Uppsala, Sweden.

Meijering, A., and D. Gianola. 1985. Observations on sire evaluation with categorical data using heteroscedatic mixed linear models. *J. Dairy Sci.* 68:1226-1232.

Misztal, I., K. Weigel, and T. J. Lawlor. 1995. Approximation of estimates of (co)variance components with multiple-trait restricted maximum likelihood by multiple diagonalization for more than one random effect. *J. Dairy Sci.* 78:1862-1872.

SAS Institute Inc. 1988. SAS/STAT® User's Guide, Release 6.03 Edition. SAS Inst. Inc., Cary, NC.

Strandberg, E. and L. Hakansson. 1994. Effect of culling on the estimates of genetic correlation between milk yield and length of productive life in dairy cattle. *Proc. 5th World Cong. Genet. Appl. Livest. Prod.*, Guelph, Canada 17:77-80.

VanRaden, P. M., and E.J.H. Klaaskate. 1993. Genetic evaluation of length of productive life including predicted longevity of live cows. *J. Dairy Sci.* 76:2758-2764.

VanRaden, P. M., and G. R. Wiggans. 1995. Productive life evaluations: Calculation,

accuracy and economic value. *J. Dairy Sci.* 78:631-638.

VanRaden, P. M., G. R. Wiggans and C.A. Ernst. 1991. Expansion of projected lactation yield to stabilize genetic variance. *J. Dairy Sci.* 74:4344-4349.

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

Vollema, A. R., and A. F. Groen. 1998. A comparison of breeding value predictors for longevity using a linear model and survival analysis. *J. Dairy Sci.* 81:3315-3320.

Vukasinovic, N, J. Moll, and N. Kunzi. 1997. Analysis of productive life in Swiss brown cattle. *J. Dairy Sci*. 80:2372-2579.

Weller, J. I. 1988. Inclusion of partial lactations in genetic analysis of yield traits by differential weighting of records. *J. Dairy Sci.* 71:1873-1879.

Weller, J. I., and E. Ezra. 1997. Genetic analysis of somatic cell concentration and female fertility of Israeli Holsteins by the individual animal model. *J. Dairy Sci.* 80:586-594.

Table1:. Number of levels of effects included in the three data sets that were analyzed.

	Data set		
Effects	1	2	3
Cows	85,965	75,825	370,406
Records	51,888	45,300	284,541
Sires	393	379	965
Herd year-season	3319	3074	16,816
Genetic groups	2	25	42

Table 2:. Estimated effects for levels of the class effects on remaining days of herd life.

	Effect							
	Parity						Pregnancy status	
Level	1	2	3	4	5	6	open	pregnant
Effect (d)	62	-56	-108	-123	-51	0	0	420

Table 3: Regression coefficients, coefficients of determination (R^2), and the correlations between actual and predicted function of variance components based on the nonlinear analysis model

Dependent variable[1]	Coefficients[2]		R^2	Correlation
	b0	b1		
RG	0.95	293	0.93	0.96
EHLX	1081	224,010	0.98	0.99
Square root of residual variance of PHL	656	45,256	0.86	0.93

[1]The dependent variables are explained in the text.
[2]The analysis model was: $y = b0 + (b1/CHL_i) + e_i$ where y is the dependent variable, and CHL is days for first calving to cut date

Table 4: Regression coefficients and correlations between estimated sire breeding values of HL and PHL with and without inclusion of weighting factors.

Weighting factors	Birth year of sires	All bulls				bulls with >10 daughters			
		No. sires	Intercept	Slope	r	No. sires	Intercept	Slope	r
With	All	379	12.68	0.98	0.94	212	10.39	0.99	0.87
	>1981	157	9.54	0.99	0.86	145	10.09	0.99	0.87
	>1982	97	12.31	0.95	0.77	86	13.41	0.94	0.76
	>1983	78	14.11	0.95	0.73	68	15.70	0.94	0.72
	>1984	39	3.95	0.85	0.61	31	5.50	0.84	0.61
Without	All	379	4.85	1.31	0.93	212	3.46	1.30	0.86
	>1981	157	0.59	1.30	0.86	145	1.70	1.29	0.86
	>1982	97	-5.01	1.47	0.76	86	-3.87	1.46	0.75
	>1983	78	-5.22	1.55	0.72	68	-3.94	1.54	0.71
	>1984	39	-23.62	1.75	0.63	31	-24.79	1.73	0.62

Table 5: Correlations between EBV for herd life and milk and protein production based on genetic evaluation of the complete Israeli Holstein population.

Birth Year	No. of bulls	Correlation with herd life	
		Milk	Protein
1986	40	0.65	0.67
1987	42	0.72	0.65
1988	53	0.43	0.44
1989	33	0.54	0.46
1990	31	0.48	0.64
1991	40	0.59	0.60
1992	42	0.63	0.65
1993	52	0.55	0.68
Total	340	0.53	0.57