

# GENETICS AND BREEDING

## Use of Linear Type and Production Data to Supplement Early Predicted Transmitting Abilities for Productive Life

K. A. WEIGEL,\* T. J. LAWLOR, JR.,† P. M. VANRADEN,‡  
and G. R. WIGGANS‡

\*Department of Dairy Science,  
University of Wisconsin, Madison 53706

†Holstein Association USA, Inc.,  
Brattleboro, VT 05301

‡Animal Improvement Programs Laboratory,  
Agricultural Research Service, USDA,  
Beltsville, MD 20705-2350

### ABSTRACT

Genetic evaluations for the length of productive life based on actual DHIA culling data have been available in the US since January 1994. Although substantial genetic variation in productive life exists, the reliability of selection is often low, particularly for recently progeny-tested bulls having daughters that have not yet had an opportunity to be culled. Correlated production and conformation traits, which have higher heritability than productive life and are available earlier in life, may be used to enhance evaluations of productive life for young bulls that have little or no direct culling information available. Genetic correlations between productive life and milk, fat, dairy form, and udder traits ranged from +0.22 to +0.46. The maximum reliability of the indirect prediction of productive life from 16 correlated type and production traits was 0.56, and the maximum reliability from a subset of 10 traits was 0.51. Indirect information about productive life that was derived from type and production traits was combined with actual culling information to increase the total amount of available information for many recently progeny-tested bulls. The procedures described herein for enhancing direct evaluations for the productive life of dairy sires with indirect information about production and type were implemented by the USDA Animal Improvement Programs Laboratory and the Holstein Association USA in July 1994.

(**Key words:** productive life, herd life, conformation, indirect prediction)

**Abbreviation key:** **PL** = productive life, **REL** = reliability.

### INTRODUCTION

The productive life (**PL**) of a cow, as measured in USDA genetic evaluations, refers to the length of time between first calving and culling from the herd (12). The maximum value of PL is 84 mo, and credit is given for up to 10 mo of PL per lactation (12). No adjustment is made for level of milk production, so PL, which is also commonly referred to as true herd life, reflects both voluntary and involuntary culling. Genetic selection for increased PL is expected to result in improved general health, production, and reproduction, which can increase dairy farm profitability by decreasing the number of replacement heifers needed, by allowing rearing costs to be dispersed over a longer period, and by increasing the number of cows producing at a mature level (1, 10).

Extensive culling data of progeny are not available until relatively late in the life of a dairy sire, but VanRaden and Klaaskate (12) have developed a procedure for projecting PL observations of cows that were still alive at the time of evaluation. Completed and projected data for PL are combined with extensive information on relatives to obtain useful predictions of genetic merit for PL earlier in the life of a dairy sire (12). National genetic evaluations for PL based on DHIA culling data were implemented in the US in January 1994 (12, 15). However, the heritability of PL is low, particularly for projected records, and many recently progeny-tested bulls have few progeny with actual culling data; PL evaluations of these progeny-tested bulls are largely based on information about the parents. For these reasons, reliability (**REL**) of PTA for PL is low for many important bulls.

Several conformation traits are correlated with PL and involuntary culling (3, 4, 5, 8, 9, 11). In addition, milk production traits are highly correlated with PL because of their influence on voluntary culling decisions. Prior to the development of direct genetic

Received April 14, 1997.

Accepted February 19, 1998.

evaluations for PL, genetic selection for PL occurred only through selection for improved conformation.

Henderson (7) presented a method for obtaining indirect BLUP predictions of breeding values for missing traits using multiple-trait BLUP. Boldman et al. (3) used the method of Henderson (7) to obtain sire model BLUP predictions of genetic merit for PL of Holstein bulls from PTA for linear type traits. Indirect predictions of PL are available earlier in life than are direct predictions, so indirect predictions may contain valuable information for many young bulls (3, 6, 16).

The objective of this study was to develop a simple procedure to combine direct PL information from DHIA culling data with indirect PL information from correlated production and type traits to obtain more accurate early predictions of genetic merit for PL in dairy bulls.

**MATERIALS AND METHODS**

**Data for Parameter Estimation**

Data used in the current study were a subset of those used by Short and Lawlor (11) and included daughters of progeny-tested bulls enrolled in the Holstein Association Sire Evaluation for Type Program between 1983 to 1985 and their herdmates. The cows chosen were those born before 1983 because they had the opportunity to live to 7 yr of age before the Short and Lawlor (11) study began. Therefore, all cows used for parameter estimation had data from actual observations for the dependent variable of interest (months of lactation at 84 mo of age). This situation was in contrast to routine USDA genetic evaluations for PL, which also included projected PL for cows that were still alive at the time of analysis. Data from 125,887 registered (63% of total) and grade (37% of total) daughters of 1677 Holstein sires were included in this study. Although different estimates for heritability and genetic correlations can sometimes be obtained from registered and grade cows, the use of a combined data set of registered and grade cows in this analysis was necessary for the following reason. The weights for indirect prediction of PL, which were derived in this study, were applied to the PTA of the progeny-tested bulls, which were sampled in registered, grade, and mixed herds. Therefore, it would have been inappropriate to estimate genetic parameters using a subset containing only registered or grade cows. Short and Lawlor (11) defined true herd life as the time from first calving to culling; maximum age was 84 mo, and no limit was placed on lactation length. The estimated genetic correlation between true herd life and PL as defined by the USDA (which includes a maximum credit of 10 mo

per lactation) was 0.984. Genetic correlations between PL and milk, fat, and 14 linear type traits were estimated using multiple-trait REML in a sire model. First lactation milk and fat data were used, but protein data were not used because they were not available for all cows. Linear type classification scores that were obtained at an age nearest 30 mo were used.

**Full Model for Indirect Prediction**

An indirect estimate of PTA for PL ( $\hat{u}_{ind}$ ) was obtained from PTA for milk, fat, and 14 linear type traits using the following equation (3).

$$\hat{u}_{ind} = Cov[u_{PL}, \mathbf{u}] [Var(\mathbf{u})]^{-1} \hat{\mathbf{u}}$$

where  $u_{PL}$  = transmitting ability for PL,  $\mathbf{u}$  = vector of true transmitting abilities for production and type traits, and  $\hat{\mathbf{u}}$  = vector of multiple-trait BLUP predictions of  $\mathbf{u}$ . The method is approximate in the current situation because PTA for linear type traits are calculated using a multiple-trait BLUP, but PTA for milk and fat are currently calculated using a single-trait BLUP. The REL of  $\hat{u}_{ind}$  was calculated using the following expression:

$$REL_{ind} = \frac{Cov[u_{PL}, \mathbf{u}] [Var(\mathbf{u})]^{-1} [Var(\hat{\mathbf{u}})]}{[Var(\mathbf{u})]^{-1} Cov[u_{PL}, \mathbf{u}] / Var(u_{PL})}$$

and

$$\max(REL_{ind}) = \frac{Cov[u_{PL}, \mathbf{u}] [Var(\mathbf{u})]^{-1} Cov[u_{PL}, \mathbf{u}]}{Var(u_{PL})},$$

which occurs when  $Var(\hat{\mathbf{u}}) = Var(\mathbf{u})$  (i.e., with an infinite amount of data on traits in  $\mathbf{u}$ ).

Direct and indirect PL predictions were combined in a weighted mean as follows;

$$\hat{u}_{comb} = w_{dir} \hat{u}_{dir} + w_{ind} \hat{u}_{ind}$$

where

$$w_{dir} = \frac{(1 - REL_{ind} \times c)}{(1 - REL_{ind} REL_{dir} \times c^2)},$$

$$w_{ind} = \frac{(1 - REL_{dir} \times c)}{(1 - REL_{ind} REL_{dir} \times c^2)},$$

$$c = 1 + \frac{[DE_{both} / DE_{dir} DE_{ind}]}{\sqrt{[(4 - h_{dir}^2) (4 - h_{ind}^2) / (h_{dir}^2 h_{ind}^2)]}}$$

and DE = daughter equivalent (14). Note that  $c$ , which is a measure of the lack of independence be-

tween direct and indirect evaluations, is a function of direct and indirect trait heritabilities and of the proportion of progeny evaluated for type and production traits that also have direct culling data available. The quantity  $c$  can be derived as  $1 + [\text{Cov}(e_{\text{dir}}, e_{\text{ind}}) / \text{Cov}(u_{\text{dir}}, u_{\text{ind}})] (\text{DE}_{\text{both}} / \text{DE}_{\text{dir}} \text{DE}_{\text{ind}})$ . The covariance of direct and indirect daughter means equals the genetic covariance multiplied by  $(\text{DE}_{\text{both}} / \text{DE}_{\text{dir}} \text{DE}_{\text{ind}})$ , and we assumed that direct and indirect daughter means are regressed toward the parent average by  $\text{REL}_{\text{dir}}$  and  $\text{REL}_{\text{ind}} / \max(\text{REL}_{\text{ind}})$ , respectively. Only daughters with both direct and indirect observations contribute to the error covariance. The  $c$  term represents the covariance of the daughter means divided by  $\text{Cov}(u_{\text{dir}}, u_{\text{ind}})$ . If genetic and phenotypic correlations are assumed to be equal, then  $[\text{Cov}(e_{\text{dir}}, e_{\text{ind}}) / \text{Cov}(u_{\text{dir}}, u_{\text{ind}})]$  equals the square root of  $[(4 - h_{\text{dir}}^2)(4 - h_{\text{ind}}^2) / (h_{\text{dir}}^2 h_{\text{ind}}^2)]$ , which is equal to 26.3 for direct and indirect trait heritabilities of 0.085 and 0.25, respectively. The weights,  $w_{\text{dir}}$  and  $w_{\text{ind}}$ , are then determined as follows. We know that the covariance of a genetic effect with its BLUP predictor equals REL times genetic variance, which is also the variance of the predictor. Therefore,  $\text{Cov}(\hat{u}_{\text{dir}}, u_{\text{PL}}) = \text{Var}(\hat{u}_{\text{dir}}) = \text{REL}_{\text{dir}} \text{Var}(u_{\text{PL}})$ , and  $\text{Cov}(\hat{u}_{\text{ind}}, u_{\text{PL}}) = \text{Var}(\hat{u}_{\text{ind}}) = \text{REL}_{\text{ind}} \text{Var}(u_{\text{PL}})$ . Furthermore,  $\text{Cov}(\hat{u}_{\text{dir}}, \hat{u}_{\text{ind}}) = \text{REL}_{\text{dir}} \text{REL}_{\text{ind}} c \text{Var}(u_{\text{PL}})$ . Then,  $w = \text{Cov}[(\hat{u}_{\text{dir}}, \hat{u}_{\text{ind}}), u_{\text{PL}}] / [\text{Var}(\hat{u}_{\text{dir}}, \hat{u}_{\text{ind}})]^{-1}$ . Simple rules for inverting  $2 \times 2$  matrices allow the weights to be reexpressed as  $w_{\text{dir}} = (1 - \text{REL}_{\text{ind}} \times c) / (1 - \text{REL}_{\text{ind}} \text{REL}_{\text{dir}} \times c^2)$  and  $w_{\text{ind}} = (1 - \text{REL}_{\text{dir}} \times c) / (1 - \text{REL}_{\text{ind}} \text{REL}_{\text{dir}} \times c^2)$ . Finally, approximate REL of the weighted average was calculated as  $\text{REL}_{\text{comb}} = (\text{REL}_{\text{dir}} + \text{REL}_{\text{ind}} - 2\text{REL}_{\text{ind}} \text{REL}_{\text{dir}} \times c) / (1 - \text{REL}_{\text{ind}} \text{REL}_{\text{dir}} \times c^2)$ .

### Reduced Model for Indirect Prediction

Many traits for linear type and production are highly correlated. For example, the estimated genetic correlation between strength and body depth is +0.90, and the estimated genetic correlation between rear udder height and rear udder width is +0.88. Because the indirect linear prediction of PL from correlated traits is equivalent to multiple genetic regression (6), this procedure is susceptible to problems of multicollinearity. For example, multiple regression on two highly correlated traits can lead to a large positive weight on one trait and a large negative weight on the other even though genetic correlations of these traits with PL are similar. These weights can lead to instability of indirect predictions for certain bulls and difficulty in interpretation of weights applied to certain traits.

Principal component analysis of a correlation matrix can be used to discard highly correlated variables in multiple regression problems (2, 17). In the current study, the objective was twofold: 1) to eliminate variables that were correlated with each other and 2) to eliminate variables that were not important predictors of PL. Therefore, a weighted genetic correlation matrix was constructed such that the variance of each indirect trait was proportional to the absolute value of its genetic correlation with PL. Using the principal component procedure for discarding redundant variables (2, 17), eigenvectors and eigenvalues were calculated iteratively from this weighted correlation matrix. At each iteration, the eigenvector corresponding to the smallest eigenvalue was identified, and the trait with the largest coefficient in this eigenvector was discarded. This procedure was repeated until only 10 indirect traits remained.

The effects of implementing the procedures described herein were then determined by examining the direct, indirect, and combined PTA and REL values for 851 Holstein progeny-tested bulls that were born in 1989 and 1990.

## RESULTS AND DISCUSSION

Estimated genetic correlations between PL and milk production and linear type traits are shown in Table 1. Genetic correlations of milk and fat with PL were +0.43 and +0.46, respectively, which were larger than correlations of any linear type traits with PL. Clearly, milk and fat have a large influence on voluntary culling decisions. Among the linear type traits,

TABLE 1. Heritabilities of milk production and linear type traits which are used in national genetic evaluations and estimated genetic correlations between these traits and productive life (PL).

Trait	Heritability	Correlation
Milk	0.30	+0.43
Fat	0.30	+0.46
Stature	0.42	+0.05
Strength	0.31	-0.13
Body depth	0.37	-0.07
Dairy form	0.29	+0.41
Rump angle	0.33	+0.09
Rump width	0.26	-0.04
Rear leg set	0.21	-0.02
Foot angle	0.15	+0.07
Fore udder	0.29	+0.29
Rear udder height	0.28	+0.32
Rear udder width	0.23	+0.31
Udder cleft	0.24	+0.28
Udder depth	0.28	+0.24
Teat placement	0.26	+0.22

TABLE 2. Weights for indirect prediction of genetic merit for productive life from production and conformation traits using the full model and the reduced model.

Trait	Full model	Reduced model
Milk	+0.000491	+0.000834
Fat	+0.0146	+0.0180
Stature	-0.041	...
Strength	+1.054	-0.185
Body depth	-1.088	...
Dairy form	+0.570	+0.026
Rump angle	+0.133	+0.109
Rump width	-0.121	...
Rear leg set	-0.099	...
Foot angle	+0.066	+0.114
Fore udder	+0.277	...
Rear udder height	-0.002	+0.094
Rear udder width	-0.009	...
Udder cleft	+0.006	+0.026
Udder depth	+0.216	+0.432
Teat placement	-0.061	+0.028

the udder traits (+0.24 to +0.32), teat placement (+0.22), and dairy form (+0.41) had the largest correlations with PL, which is consistent with the results of previous studies using independent data sets (3, 4). Udder traits have an important influence on involuntary culling decisions, mainly because of their influence on susceptibility to injuries and mastitis infections. Because data from the production and type traits are available relatively early in life and because those traits are more highly heritable than culling traits, production and type traits may provide useful additional information regarding PL genetic merit for many bulls that have recently been progeny tested.

Weights for the indirect prediction of genetic merit for PL from milk production and linear type traits are presented in Table 2. In the full model, the weights used for indirect prediction of PL were difficult to interpret because they correspond to coefficients of multiple genetic regression with highly correlated traits. In the reduced model, the following six linear type traits were discarded: body depth, stature, rump width, rear leg set, rear udder width, and fore udder. Ten production and conformation traits remained for the indirect prediction of PL in the reduced model: milk, fat, strength, dairy form, rump angle, foot angle, rear udder height, udder cleft, udder depth, and teat placement. The decision to stop discarding traits was based on the inspection of the genetic correlation matrix among the remaining traits and the marginal decrease in REL of indirect prediction of PL, which would have occurred by discarding one more trait. The weights for the indirect prediction of PL in the reduced model correspond more closely to the genetic

correlations with PL because multicollinearity problems have been reduced. Importantly, however, these weights are a function of the genetic (co)variances among the traits used for indirect prediction, so conclusions regarding the relative importance of production and linear type traits should be based on the estimated genetic correlations rather than the indirect prediction weights. In addition, the prediction weights are valid only for this particular set of type and production traits, and recalculation of these weights will be necessary for any future studies that involve different traits.

The maximum REL of the indirect prediction of PL using type and production data were 0.56 with the full model and 0.51 with the reduced model; therefore, little information was lost by discarding data from the six redundant or unimportant linear traits. The indirect predictions of the genetic merit for PL were generally similar using the full and reduced models, although predictions for a few individual bulls differed by 0.6 to 1.3 mo. Each of these bulls had evaluations for strength and body depth that differed by more than 1.5 genetic SD (even though these traits are very highly correlated genetically), and application of the large positive weight for strength and the large negative weight for body depth in the full model gave indirect predictions that were inconsistent with those obtained from the reduced model. Given these results, the use of a reduced number of traits for indirect prediction of PL would seem desirable, but this change has not yet been implemented by Holstein Association USA or USDA-Animal Improvement Programs Laboratory.

In practice, as shown in Table 3, the mean indirect prediction for PL obtained from type and production traits was slightly higher than the mean direct prediction for PL obtained from the culling data. This difference occurred because these progeny-tested bulls were subjected to intense pedigree selection for

TABLE 3. Mean PTA and REL for direct, indirect, and combined predictions of productive life in July 1994 for 851 Holstein bulls born in 1989 and 1990.

	Mean	
	PTA	REL
Direct prediction	+0.83	0.50
Indirect prediction		
Full model	+1.45	0.36
Reduced model	+1.33	0.33
Combined prediction		
Full model	+1.19	0.53
Reduced model	+1.14	0.52

production and type traits (which were used for indirect prediction of PL) but little or no direct selection for PL. In addition, the genetic base for PL (defined as all cows born in 1990 with records for milk production) was slightly different from the genetic base for linear type traits (defined as all cows born in 1990 that were classified for type), which could have caused a slightly positive bias in indirect predictions of PL. Approximate REL of the combined PL evaluations was slightly higher than that of the direct PL evaluations for bulls that had little culling data available for daughters at the time of the genetic evaluation. Although the mean increase in approximate REL of recently progeny-tested bulls was only 0.03, some individual bulls increased by >0.10, and this increase was accomplished with no additional costs for data collection.

### CONCLUSIONS

The length of PL of dairy cows is positively correlated with several production and linear type traits, and data from these correlated traits are available earlier in life and are more highly heritable than is PL. Additional information provided by these correlated traits can be used to increase slightly the accuracy of PL evaluations for recently progeny-tested bulls that have little or no available information about the culling of their progeny. The procedure described herein for enhancing direct PL evaluations of dairy sires with indirect information from 16 production and type traits was implemented in national genetic evaluations by the USDA Animal Improvement Programs Laboratory and the Holstein Association USA in July 1994. VanRaden et al. (13) recently compared the accuracy of early direct and combined PTA for PL using Holstein sires that obtained second-crop AI progeny between July 1994 and January 1997. The combined PTA for PL, which were calculated from first-crop progeny information using the procedures described herein, were a better predictor of the eventual PTA for PL of these bulls than were the direct first-crop predictions.

The weighted mean procedure for combining direct PL information from culling with indirect information for type and production is an approximation, and a true multiple-trait analysis of production, type, and PL would be preferred when such an evaluation becomes available. In addition, large-scale genetic

evaluations of dairy cattle using survival analysis procedures (which correctly account for censoring of the data) are rapidly becoming computationally feasible, and such procedures likely can further improve the accuracy of genetic evaluations for PL (5).

### REFERENCES

- Allaire, F. R., and J. P. Gibson. 1992. Genetic value of herd life adjusted for milk production. *J. Dairy Sci.* 75:1349–1356.
- Beale, E.M.L., M. G. Kendall, and D. W. Mann. 1967. The discarding of variables to multivariate analysis. *Biometrika* 54: 357–366.
- Boldman, K. G., A. E. Freeman, B. L. Harris, and A. L. Kuck. 1992. Prediction of sire transmitting abilities for herd life from transmitting abilities for linear type traits. *J. Dairy Sci.* 75: 552–563.
- Brotherstone, S., and W. G. Hill. 1990. Dairy herd life in relation to type traits and production. *Proc. 4th World Congr. Genet. Appl. Livest. Prod., Edinburgh, Scotland* 14:209–212.
- Ducrocq, V. P., and J. Solkner. 1994. "The survival kit", a FORTRAN package for the analysis of survival data. *Proc. 5th World Congr. Genet. Appl. Livest. Prod., Guelph, ON, Canada* 22:51–52.
- Harris, B. L., A. E. Freeman, and E. Metzger. 1992. Analysis of herd life in Guernsey dairy cattle. *J. Dairy Sci.* 75:2008–2016.
- Henderson, C. R. 1984. *Applications of Linear Models in Animal Breeding*. Univ. Guelph, Guelph, ON, Canada.
- Rogers, G. W., M. R. Dentine, I. Miedema, and H. D. Norman. 1990. Relationships among non-production culling and linear type traits in Holsteins. *Proc. 4th World Congr. Genet. Appl. Livest. Prod., Edinburgh, Scotland* 14:213–216.
- Rogers, G. W., B. T. McDaniel, M. R. Dentine, and D. A. Funk. 1989. Genetic correlations between survival and linear type traits measured in first lactation. *J. Dairy Sci.* 72:523–527.
- Rogers, G. W., J.A.M. Van Arendonk, and B. T. McDaniel. 1988. Influence of involuntary culling on optimum culling rates and annualized net revenue. *J. Dairy Sci.* 71:3463–3469.
- Short, T. H., and T. J. Lawlor. 1992. Genetic parameters of conformation traits, milk yield, and herd life in Holsteins. *J. Dairy Sci.* 75:1987–1998.
- 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., R. J. Starckenburg, and T. J. Lawlor, Jr. 1997. Stability of genetic evaluations for productive life, somatic cell score, and yield over time. *J. Dairy Sci.* 80(Suppl. 1): 253.(Abstr.)
- VanRaden, P. M., and G. R. Wiggans. 1991. Derivation, calculation, and use of national animal model information. *J. Dairy Sci.* 74:2737–2746.
- VanRaden, P. M., and G. R. Wiggans. 1993. Genetic evaluation of herd life with an animal model. *J. Dairy Sci.* 76(Suppl. 1): 251.(Abstr.)
- Weigel, D. J., and B. G. Cassell. 1994. Differences in multiple trait prediction of transmitting abilities for herd life due to data source. *Proc. 5th World Congr. Genet. Appl. Livest. Prod., Guelph, ON, Canada* 17:73–76.
- Weigel, K. A., A. E. Freeman, M. E. Kehrli, Jr., J. R. Thurston, and D. H. Kelley. 1992. Relationship of in vitro immune function with health and production in Holstein cattle. *J. Dairy Sci.* 75:1672–1679.