Use of Correlated Trait Information to Improve the Accuracy of Early Predictions of Breeding Value for Length of Productive Life

Kent A. Weigel

Department of Dairy Science, University of Wisconsin, Madison, 53706, USA

ABSTRACT

Procedures are presented for approximate multiple-trait prediction of productive life using correlated type and production trait information to supplement culling data. Production and type traits are more highly heritable than productive life, and these traits are available earlier in life than culling data. Genetic evaluations for milk yield, fat yield, and 14 linear type traits are combined with genetic evaluation or parent average information for direct productive life (from culling data) using an approximate approach based on a weighted average and using procedures for international sire evaluations. The maximum indirect reliability of productive life calculated from 16 type and production traits was .56; however, several of the linear type traits were redundant or unimportant with respect to productive life, and it was possible to obtain indirect reliability of .51 using only 10 traits. Use of a reduced number of indirect traits can eliminate problems due to multicollinearity of highly correlated type traits and can make indirect prediction weights easier to interpret. The contribution of indirect type and production traits to productive life was most important for new progeny test bulls whose progeny had not yet been subject to culling decisions.

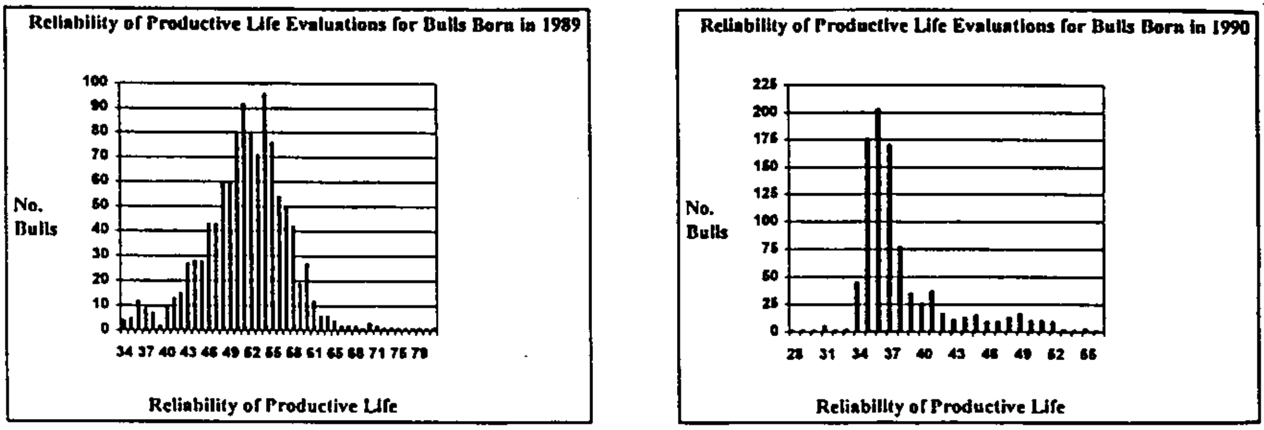
INTRODUCTION

Length of productive life (PL) is measured in the USA as the number of months in milk from first calving to 84 months of age, with a maximum credit of ten months per lactation (VanRaden and Klaaskate, 1993). Productive life is not adjusted for level of milk production, therefore it is a measure of both voluntary and involuntary culling. National genetic evaluations, which have been available since January 1994, use both completed records (culled cows and cows > 84 months of age) and records in progress (cows greater than 36 months of age which are still alive at time of evaluation). Because heritability of PL is only .085, reliability (REL) of PL calculated from culling data is low for many recent progeny test bulls (Figure 1). Nearly all bulls born in 1989 have daughters ≥ 36 months of age with PL records in progress, but most bulls born in 1990 have only parent average information for PL, even though they already have official evaluations for production and type traits.

An alternative to direct genetic evaluation of culling data is indirect prediction of genetic merit for PL from correlated production and type traits. Production and type trait data are available earlier in life than culling data, and these traits are more highly heritable than are direct measures of PL. Boldman et al. (1992) applied the techniques developed by Henderson (1984) for indirect multiple-trait BLUP prediction of breeding values for missing traits to the situation of indirect prediction of breeding values for herd life from breeding values of linear type traits.

While direct PL evaluations from culling data and indirect PL evaluations from correlated traits each provide useful information, the optimal approach would be to use both direct and indirect information. This could be accomplished by multiple-trait BLUP evaluation of PL, production, and linear type traits. However, multiple-trait procedures for simultaneous evaluation of PL, production, and type traits have not yet been implemented in the USA, because PL data arrive later than type and production data, and because not all progeny have type classification data. The objective of this study was to evaluate several possible methods for combining or blending direct and indirect PL information into approximate multiple-trait evaluations which would be of higher accuracy than direct PL evaluations for recent progeny test bulls.

Figure 1. Reliability of direct productive life evaluations for US Holstein bulls born in 1989 and 1990.



MATERIALS AND METHODS

Data used in Parameter Estimation

Data used in the present study were a subset of those used by Short and Lawlor (1993) and consisted of daughters of progeny test bulls enrolled in the Holstein Association Sire Evaluation for Type Program from 1983 to 1985 and their herdmates. Cows born before 1983 were chosen, because these cows had an opportunity to live to 7 years of age before the Short and Lawlor (1993) study began. Short and Lawlor (1993) defined PL as the time from first calving to 84 months of age, with no limit on lactation length. The genetic correlation between this measure of PL and the measure used by USDA in national genetic evaluations, which includes a maximum of 10 months per lactation, was .984. Data from 125,887 registered (63% of total) and grade (37% of total) daughters of 1677 Holstein sires were included in this study. Genetic correlations among PL, milk and fat yield, and 14 linear type traits were estimated using multiple-trait REML in a sire model.

Full Model for Indirect Prediction

An indirect prediction of breeding value for PL was obtained from breeding values for correlated production and type traits as follows:

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

where u_{PL} is breeding value for PL, $\mathbf{u}$ is a vector of breeding values for correlated traits and $\hat{\mathbf{u}}$ is its multiple-trait BLUP prediction. Reliability of $\hat{u}_{indirect}$ was calculated using the following expression:

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

where

$$\mathrm{Max}\,(\,REL_{indirect}\,) = \mathrm{Cov}\,[\,u_{PL},\,\mathbf{u}\,]'\,[\,\mathrm{Var}\,(\mathbf{u})\,]^{-1}\,\mathrm{Cov}\,[\,u_{PL},\,\mathbf{u}\,]\,/\,\mathrm{Var}\,(\,u_{PL}\,)$$

Direct and indirect PL predictions were combined as follows:

$$\hat{u}_{combined} = w_{direct}\,\hat{u}_{direct} + w_{indirect}\,\hat{u}_{indirect}$$

where
$w_{direct} = (1 - REL_{indirect} * c) / (1 - REL_{indirect}\,REL_{direct} * c^2)$, $w_{indirect} = (1 - REL_{direct} * c) / (1 - REL_{indirect}\,REL_{direct} * c^2)$ and $c = 1 + [DE_{both} / DE_{direct}\,DE_{indirect}] * \mathrm{Sqrt}[(4 - h^2_{direct})(4 - h^2_{indirect}) / (h^2_{direct}\,h^2_{indirect})]$

is a function of direct and indirect trait heritabilities and the proportion of progeny evaluated for type and production traits which also have direct culling data available (DE = daughter equivalent). Further,

$$REL_{combined} = (REL_{direct} + REL_{indirect} - 2\,REL_{indirect}\,REL_{direct} * c) / (1 - REL_{indirect}\,REL_{direct} * c^2)$$

Reduced Model for Indirect Prediction

Many linear type and production traits are highly correlated. For example, the r_g between strength and body depth is .90, and the r_g between udder height and udder width is .88. Because indirect linear prediction of PL from correlated traits is equivalent to multiple genetic regression (Harris et al., 1992), it 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 the genetic correlations between these traits and PL are similar. This can potentially lead to instability of indirect predictions for some bulls and difficulty in interpretation of weights applied to certain traits.

Principal component analysis of a correlation matrix can be used to discard redundant variables in multiple regression problems (e.g., Beale et al., 1967; Weigel et al., 1992). In the current study, however, it was desirable to eliminate both redundant variables and variables which are unimportant with respect to their relationship with 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 productive life. Using the principal component procedure for discarding redundant variables, eigenvectors and eigenvalues were calculated from this weighted correlation matrix in an iterative manner. 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 ten indirect traits remained.

MACE Analysis of PL and Correlated Traits

The multiple-trait across country evaluation (MACE) procedure was developed as a method to combine sire evaluation data from multiple countries when genetic correlations among countries are < 1 (Schaeffer, 1994).

The situation of international sire evaluation is analogous to the current problem: sire evaluation data for many correlated traits from different sources are to be combined in an optimal manner. For this reason, MACE procedures were used to simultaneously analyze PL data and data for the ten correlated production and linear type traits selected by the principal component procedure described in the previous section. All bulls with USA genetic evaluations for PL, production, or linear type traits were used in calculating variance of sire transmitting ability using the single-trait procedure described by Schaeffer (1995). For computational reasons, data used in the MACE analysis were limited to the most informative bulls. Bulls which were non-AI proven were discarded if they had daughters in less than 150 herds. Bulls were required to have at least 10 daughters in 10 herds with productive life data, 40 daughters in 30 herds with production data, or 30 daughters in 20 herds with type data for inclusion in the MACE analysis. Furthermore, bulls were required to have ≥ 90% US daughters (i.e., Canadian bulls were excluded) and a USA sire and maternal grandsire.

Productive life data were limited to bulls b from 1975 to present, and production and lin type data were limited to bulls born from 19 to present. The primary contribution correlated production and linear type da occurs for bulls which do not yet have direct productive life evaluations based on a large number of second-crop AI daughters, so it was expected that little information would be lost by excluding indirect trait information for bulls born prior to 1983. A total of 6553 bulls were included in the MACE analysis; 6304 with productive life data, 4358 with production data and 3764 with type data. The pedigree file included 7026 total bulls and 44 genetic groups assigned by year of birth and inheritance pathway.

RESULTS

The following production and linear type traits were used to calculate an indirect prediction of genetic merit for PL in the full model using the multiple genetic regression procedure described earlier:

Table 1. Heritabilities of indirect type and production traits and their genetic correlations with productive life.

Trait	h^2	r_g with PL	Trait	h^2	r_g with PL
Milk	.25	+.43	Rear Leg Set	.21	-.02
Fat	.25	+.46	Foot Angle	.15	+.07
Stature	.42	+.05	Fore Udder	.29	+.29
Strength	.31	-.13	Udder Height	.28	+.32
Body Depth	.37	-.07	Udder Width	.23	+.31
Dairy Form	.29	+.41	Udder Cleft	.24	+.28
Rump Angle	.33	+.09	Udder Depth	.28	+.24
Rump Width	.26	-.04	Teat Placement	.26	+.22

In the reduced model, six linear type traits were discarded, in the following order: body depth, stature, rump width, rear leg set, udder width and fore udder. Ten correlated production and type traits remained for indirect prediction of PL: milk yield, fat yield, strength, dairy form, rump angle, foot angle, udder height, udder cleft, udder depth and teat placement. These traits were used to calculate an indirect prediction of PL for each bull using the same procedure as used in the full model. Weights used for indirect prediction of transmitting ability for PL in the full and reduced models are shown below.

Table 2. Weights for indirect prediction of productive life using the full model and the reduced model.

Trait	Model Full	Reduced	Trait	Model Full	Reduced
Milk	+.000491	+.000834	Rear Leg Set	-.099	.....
Fat	+.0146	+.0180	Foot Angle	+.066	+.114
Stature	-.041	.....	Fore Udder	+.277	.....
Strength	+1.054	-.185	Udder Height	-.002	+.094
Body Depth	-1.088	.....	Udder Width	-.009	.....
Dairy Form	+.570	+.026	Udder Cleft	+.006	+.026
Rump Angle	+.133	+.109	Udder Depth	+.216	+.432
Rump Width	-.121	.....	Teat Placement	-.061	+.028

Maximum REL of indirect prediction was .56 with the full model and .51 with the reduced model, and these values were consistent with literature estimates (e.g., Boldman et al., 1992). The main differences between weights for indirect prediction derived from the full model and those derived from the reduced model were in the sign and magnitude of the weights for highly correlated traits. In the reduced model, the sign associated with each weight corresponded to the direction of the genetic correlation with PL. In the full model, however, stature, udder height, udder width, and teat placement had negative weights, even though these traits were positively correlated with PL. Similarly, strength received a very large positive weight, despite its negative correlation with PL. In fact, the opposing weights given to strength and body depth were far larger than those assigned to other traits which were more closely related to PL, such as udder traits. Linear type traits are evaluated using multiple-trait BLUP procedures in the USA, so it is unlikely that a particular bull would have genetic evaluations for highly correlated traits, such as strength and body depth, which differ widely. However, a few individual bulls were identified for prediction of PL using multiple linear regression with these highly correlated traits could be unstable:

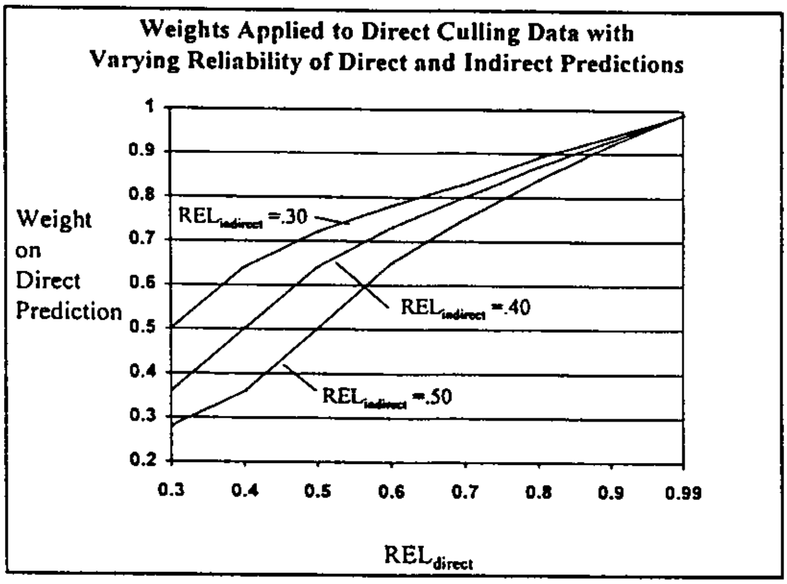
Table 3. Potential problem bulls with respect to indirect prediction of productive life PTA using the full model.

Bull	Linear type PTA Strength	Body depth	Indirect PL PTA Full model	Reduced model
2140297	+2.40	+5.04	-0.31	+0.96
2019608	+3.78	+2.32	+0.16	-0.49
2089579	-0.59	+0.87	+1.98	+1.41

For these bulls, strength and body depth evaluations differed considerably, so application of a large positive weight to strength PTA and a large negative weight to body depth PTA gave results which were inconsistent with those obtained with the reduced model.

Weights applied to indirect and direct predictions of PTA PL are shown in the following graph as a function of REL of each of the predictions:

Figure 2. Relative weights applied to direct culling data and indirect type and production data in approximate multi-trait prediction of productive life.



The estimated variances of sire transmitting ability for productive life and for correlated production and type traits calculated for use in indirect prediction and in the MACE analysis are shown below:

Table 4. Estimated variance of sire transmitting ability for direct and indirect traits.

Trait	Var (TA)	Trait	Var (TA)
Productive Life	2.38	Rear Leg Set	2.00
Milk	534000	Foot Angle	1.48
Fat	692	Fore Udder	3.05
Stature	3.72	Udder Height	2.50
Strength	2.05	Udder Width	1.99
Body Depth	2.41	Udder Cleft	1.94
Dairy Form	2.56	Udder Depth	2.53
Rump Angle	3.27	Teat Placement	2.77
Rump Width	1.83		

Units of measurement are lb. for production traits, months for PL, and linear score (1-50 scale) for type tarts.

In the following table, the average direct, indirect and combined PL evaluations are shown by year of birth for bulls included in this study (analyses of results included progeny test bulls with at least 40 daughters in 30 herds for production traits and 30 daughters in 20 herds for linear type traits):

Table 5. Average direct, indirect and combined evaluations of productive life by year of birth for bulls in this study.

YOB	n	Direct	Indirect Full	Indirect Reduced	Combined Full	Combined Reduced	MACE
Bulls with ≥ 10 daughters with culling data							
1983	295	+0.02	+0.24	+0.20	+0.03	+0.01	-0.05
1984	415	+0.05	+0.17	+0.14	+0.04	+0.04	-0.04
1985	503	+0.29	+0.40	+0.35	+0.33	+0.31	+0.23
1986	545	+0.07	+0.56	+0.39	+0.19	+0.15	+0.00
1987	538	+0.08	+0.68	+0.54	+0.24	+0.20	+0.10
1988	589	+0.37	+1.11	+0.96	+0.63	+0.58	+0.45
1989	577	+0.83	+1.46	+1.30	+1.16	+1.10	+1.00
1990	87	+0.90	+1.39	+1.40	+1.23	+1.23	+1.21
Bulls with < 10 daughters with culling data							
89-90	187	+0.81	+1.45	+1.40	+1.28	+1.23	+1.10

Obviously changes were largest for young progeny test bulls which had < 10 daughters with culling data and, hence, no official PL evaluation (i.e., parent average information only). It is important to note that the average indirect PL evaluation obtained from correlated production and type traits was generally larger than the average direct evaluation obtained from culling data. This occurred because these progeny test bulls had been subjected to intense pedigree selection for production and type traits which were used for indirect prediction of PL. An additional reason for this discrepancy was that the genetic base used for PL (all milk-recorded cows born in 1990) is slightly different from the genetic base for linear type traits (all type-classified cows born in 1990); this could lead to slight upward bias in indirect predictions of PL. For bulls which have second-crop AI daughters and bulls whose first-crop daughters have completed multiple lactations, combined and MACE evaluations differed very little from direct (raw) PL evaluations. For bulls born since 1988, however, combined and MACE evaluations were substantially different from direct evaluations, particularly for bulls with only parent average information for direct PL. The MACE procedure should accurately account for genetic base differences between PL, production, and linear type traits. Therefore, one can take the difference between combined and MACE evaluations as a rough estimate of the amount of bias due to differences in the genetic base of PL and the genetic base of linear type traits. A further advantage of the MACE procedure is that it properly combines information from either single or multiple-trait evaluation models, whereas the indirect prediction method of Henderson (1984) was developed under the assumption that all correlated traits were evaluated using a multiple-trait model. Therefore, this application of indirect prediction was an approximation, because linear type traits were evaluated with a multiple-trait model, but milk and fat were each evaluated using single-trait models.

The following table shows SD of direct, indirect, combined, and MACE PTA for PL by year of birth for bulls in this study:

Table 6. SD of direct, indirect and combined evaluations of productive life by year of birth for bulls in this study.

YOB	Direct	Indirect		Combined		MACE
		Full	Reduced	Full	Reduced	
Bulls with ≥ 10 daughters with culling data						
1983	1.21	1.07	1.06	1.19	1.21	1.21
1984	1.25	1.05	1.07	1.22	1.25	1.27
1985	1.19	1.00	1.04	1.15	1.17	1.18
1986	1.16	1.03	1.02	1.12	1.15	1.17
1987	1.18	1.04	1.03	1.14	1.17	1.19
1988	1.16	1.11	1.09	1.16	1.18	1.17
1989	1.04	1.00	1.01	1.02	1.04	1.12
1990	1.04	0.82	0.82	0.98	0.99	1.12
Bulls with < 10 daughters with PL data						
1989-90	0.80	0.94	0.89	0.81	0.75	1.06

The SD of indirect prediction was slightly smaller than the SD of direct prediction, because only a portion of the genetic variation in PL was described by the production and type traits. The SD of combined evaluations were similar to SD of direct evaluations, while the SD of MACE evaluations were slightly larger than SD of direct evaluations.

The following table shows the approximate standard error of prediction (SEP = Sqrt[(1 - REL) * Var(TA)]) and SD of the difference between combined, MACE, and direct PL evaluations by year of birth for bulls in this study:

Table 7. Standard error of prediction and standard deviation of change in combined evaluations using weighted average and MACE procedures.

YOB	SEP	Combined		SD (MACE - Direct)
		SD (Full - Direct)	SD (Reduced - Direct)	
Bulls with ≥ 10 daughters with PL data				
1983	0.72	0.16	0.14	0.34
1984	0.80	0.17	0.16	0.40
1985	0.85	0.22	0.20	0.48
1986	0.90	0.26	0.23	0.52
1987	0.93	0.26	0.23	0.53
1988	0.95	0.29	0.25	0.51
1989	1.05	0.34	0.30	0.54
1990	1.11	0.33	0.33	0.58
Bulls with < 10 daughters with PL data				
1989-90	1.23	0.54	0.49	0.93

The SD of the difference between combined and direct evaluations using ten correlated traits were slightly smaller than the SD of the difference between combined and direct evaluations using the full sixteen trait model. Furthermore, the SD of the difference between MACE and direct PL evaluations were substantially larger than those corresponding to the combined evaluations. It appears that the MACE procedure places greater weight on correlated production and type traits than does the approximate combining procedure presented herein.

Average approximate REL of direct, indirect, and combined PL evaluations by year of birth are shown below for bulls in this study:

Table 8. Approximate reliability of direct, indirect, and combined productive life evaluations of bulls in this study by year of birth.

YOB	Direct	Indirect Full	Indirect Reduced	Combined Full	Combined Reduced	MACE
Bulls with ≥ 10 daughters with PL data						
1983	.78	.43	.40	.79	.79	.84
1984	.73	.42	.39	.74	.74	.81
1985	.70	.41	.38	.71	.71	.78
1986	.66	.40	.37	.67	.67	.75
1987	.64	.38	.35	.65	.65	.74
1988	.62	.38	.35	.63	.63	.74
1989	.54	.37	.34	.57	.56	.73
1990	.48	.36	.33	.51	.51	.69
Bulls with < 10 daughters with PL data						
89-90	.37	.35	.31	.42	.41	.57

Reliability of combined evaluations was slightly higher than that of direct evaluations for bulls which did not yet have a large number of direct PL observations. Although differences in REL between direct and combined evaluations were small, an average gain in REL of 3% is substantial, given that it involves no additional data collection costs. The REL of combined evaluations is an approximation, so the actual gain in information due to correlated traits may differ slightly from the figures shown above. Reliability of MACE evaluations was substantially larger than that of combined evaluations, although this is also an approximation. The reason for this difference is unknown, but this is consistent with the large SD of (MACE PTA - Direct PTA) shown earlier.

DISCUSSION

Three procedures for obtaining approximate multiple-trait PL evaluations using correlated production and type traits were described and tested herein. While it is recognized that simultaneous multiple-trait BLUP evaluation of PL, production, type and perhaps health traits (e.g., SCC) would be optimal, such approximations may useful in the short term. The multiple genetic regression procedure presented herein is easy to apply on a large scale, and it can increase the accuracy of PL evaluation for recent progeny test bulls which have limited culling data. MACE is more computationally demanding on a large scale, but it has several theoretical advantages relative to the multiple regression-based procedures. Differences in REL and in SD of the change in PL PTA using MACE and the multiple genetic regression procedures should be further investigated.

ACKNOWLEDGMENTS

Productive life and production data for the present study were generously provided by the USDA Animal Improvement Programs Laboratory, and type data were supplied by the Holstein Association USA, Inc. Financial support of this project was provided by World-Wide Sires, Inc.

REFERENCES

Beale, E. M. L., M. G. Kendall, and D. W. Mann. 1967. The discarding of variables in multivariate analysis. Biometrika 54:357.

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 of linear type traits. J. Dairy Sci. 75:552.

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

Henderson, C. R. 1984. *Applications of Linear Models in Animal Breeding*. Univ. Guelph, Guelph, ON, Canada.

Schaeffer, L. R. 1994. Multiple-country comparison of dairy sires. J. Dairy Sci. 77:2671.

Schaeffer, L. R., R. Reents, and J. Jamrozik. 1995. Factors influencing international comparisons of dairy sires. J. Dairy Sci. 78(Suppl. 1):221.

Short, T. H., and T. J. Lawlor, Jr.. 1992. Genetic parameters of conformation traits, milk yield, and herd life in Holsteins. J. Dairy Sci. 75:1987.

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.

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.