

Expansion of Projected Lactation Yield to Stabilize Genetic Variance

P. M. VanRADEN, G. R. WIGGANS, and C. A. ERNST
Animal Improvement Programs Laboratory
Agricultural Research Service, USDA
Beltsville, MD 20705-2350

ABSTRACT

Information on partial lactations often is included in genetic evaluations by predicting the cow's eventual 305-d yield. Such projected yields have less phenotypic and genetic variation than completed yields but were modeled as having greater or equal variation in evaluations. Analysis of first lactations from 48,424 daughters of 844 Holstein sires indicated that yields predicted early (46 to 75 d) in lactation had less than one-half as much additive genetic variance as completed yields. Multiple-trait REML estimates of genetic correlations of predicted and completed yields were all above .92, indicating that early lactation information is valuable if modeled appropriately. Expanded records with genetic variances equal to those of completed yields and new lactation length weights were derived. Expanded records have larger error variances than either projected or completed yields and, thus, are given less weight when included in animal model evaluations. Genetic gains are expected to increase only .2 to .3%, but more stable genetic evaluations should result from use of expanded records, particularly for animals evaluated primarily from first lactation records in progress. (Key words: projected yield, genetic evaluation, heterogeneous variance)

Abbreviation key: RIP = record in progress.

INTRODUCTION

Genetic evaluations often include projected (predicted) yields computed from information on partial lactations [e.g., (5, 9)] if completed

yield information is unavailable for a lactation. Projection factors may make use of several variables such as cumulative yield, last sample-day yield, and herd average to predict a cow's eventual 305-d yield as accurately as possible (7, 9). Although intuitively appealing, the projected yields produced by this strategy are not optimal for use in genetic evaluations. Projected yields often are treated as having error variances greater than those of completed yields so that they will receive less emphasis in evaluations (2, 6). Projected yields actually have less variance than completed yields because coefficients of determination (squared correlations of predicted with true yield) are >1 .

Weller (5) found that sire and error variances of projected yields were much less than corresponding variances of completed yields, especially for projections made early in lactation. He proposed adjusting coefficients of mixed model equations to make assumptions of the model match actual distribution of records. Two adjustments were needed: 1) elements of the sire coefficient matrix were no longer simply 0 or 1 but, instead, equaled regressions of sire effect in projected records on sire effect in completed records for daughters represented by projected records and 2) diagonals of the error variance matrix reflected the different error variances of projected and completed records.

Projected yields have received less weight than completed yields in USDA animal model (6) and Modified Contemporary Comparison (2) evaluations, but reduced genetic variances were ignored. Permanent environmental and herd-sire interaction effects also may have reduced variance in projected records, but estimates of these reductions have not been reported. Weighting factors for error variance (lactation length weights) were correlations (2) or squared correlations (6) from the projection procedure (9) rather than ratios of error variances (5). Thus, improved modeling of projected yields was desired, but sire model

Received May 10, 1991.
Accepted July 25, 1991.

techniques of Weller (5) could not be applied directly.

The objectives of this study were 1) to develop procedures to account for unequal variances of projected and completed yields in an animal model, 2) to estimate variances and covariances for projected and completed 305-d standardized yields, and 3) to investigate effects of adjusting for these unequal variances in routine genetic evaluations.

MATERIALS AND METHODS

For genetic evaluations, a cow's genes usually are modeled as having the same effect on projected and completed records. Because partial lactation records usually are less heritable than completed records (1, 5), projected yields must have larger phenotypic variance than completed yields if genetic variance is to be equal. Projected yields actually have less phenotypic variance than completed yields because predictions have less variance than the variables that they predict. Thus, projected yields have variances less than assumed in evaluation models. Expansion of projected yields to increase genetic variance may be the simplest way to make data and model agree.

Expansion Theory

Let t represent a cow's total 305-d yield and p the best projection (prediction) of t from partial lactation information. Regression of t on p is 1, and p is chosen to have the largest possible correlation with t . Then, t is the sum of p plus an error of projection ($t - p$), or

$$t = p + (t - p).$$

Because p is the expected value of t given the partial lactation information, the error of projection is uncorrelated with p so that

$$\begin{aligned} \text{Var}(t) &= \text{Var}(p) + \text{Var}(t - p); \\ \text{Cov}(p, t) &= \text{Var}(p). \end{aligned}$$

If p does not predict t perfectly, $\text{Var}(t - p)$ is greater than 0, and $\text{Var}(p)$ is less than $\text{Var}(t)$. Expected value of $(t - p)$ is 0, which implies that

$$E(p) = E(t) = m$$

where m represents a management group effect or some other mean.

Because $\text{Var}(p)$ is less than $\text{Var}(t)$, variances of effects included in p likely are less than variances of corresponding effects in t . Evaluation models can account for these reduced variances by either including regressions of p on effects in t as described by Weller (5), treating records of different lengths as separate traits in a multiple-trait analysis, or expanding variance of p until genetic variances are equal.

Let q be an expanded record obtained by multiplying the deviation of p from m by x (where x is an expansion factor) and adding back m , or

$$q = x(p - m) + m.$$

In practice, m is not known, and a management group mean or other estimate of m (\hat{m}) must be substituted. Genetic rankings are affected little by the method used to obtain \hat{m} if management groups consist primarily of records of the same length. Expected values of q would differ from those of p and t by $(x - 1)(m - \hat{m})$. For x close to 1, values of \hat{m} have little effect. If x is identical for all cows in a management group, the constant $(x - 1)(m - \hat{m})$ is included in animal model estimates of management group effects but does not affect other predictions.

Expansion Factors

Expansion factors can be calculated in two ways, using either phenotypic correlations or genetic standard deviations. For this study, theoretical expansion factors used only phenotypic information and were computed as reciprocals of previous lactation length weights presented by Wiggans et al. (6), which are squared correlations, $[\text{Corr}(p, t)]^2$. Actual expansion factors for each trait were computed as standard deviations of sire effects in t divided by standard deviations of sire effects in p .

Theoretical expansion factors cause the direction of best prediction to be reversed such that the total yield t is the best predictor of the expanded record q , or $E(q|t) = t$. Theoretical expansion factors are chosen as

$$x = \text{Var}(t)/\text{Var}(p)$$

$$= 1/[\text{Corr}(p,t)]^2.$$

Variance of q then is $x^2[\text{Var}(p)] = x[\text{Var}(t)]$, covariance of q with t is $x[\text{Cov}(p,t)] = \text{Var}(t)$, and variance of $(q - t)$ is $(x - 1)[\text{Var}(t)]$. These equalities imply that q contains t plus an error $(q - t)$ uncorrelated with t or

$$q = t + (q - t);$$

$$\text{Var}(q) = \text{Var}(t) + \text{Var}(q - t).$$

Thus, q contains exactly the same effects as t plus an additional, independent error. Expected values of q and t are equal provided that m equals m . Therefore, inclusion of expanded yields as data for genetic evaluations is simple and does not require altering coefficient matrices as in Weller (5).

The error $q - t$ is uncorrelated with t but could be correlated with $q - t$ of other cows, and genetic and other effects expressed in q may not actually be the same as those in t . Use of theoretical x assumes that q (and p) has a heritability of h^2/x , where h^2 is heritability of t . Genetic correlation of q with t (p with t) is assumed to be 1. Actual heritabilities and genetic correlations may differ from these assumed values, and some caution is advised if using theoretical x .

Actual expansion factors equate genetic variance of q with genetic variance of t , with the assumption that variances of all other random effects except error also will be equalized. Equal genetic variance does not guarantee that genetic effects in q and t are identical unless genetic correlation of q and t is 1. If genetic correlation with t is much less than 1, q should be excluded from data or treated as a separate trait.

Length Weights

Lactation length weights (w_{len}) should be error variance of t divided by error variance of q . Because permanent environmental effects are included in the USDA animal model, w_{len} then is ratio of variances only for temporary environmental effects (e). Variance of e for a standard length record (e_t) is

$$\text{Var}(e_t) = (1 - r)[\text{Var}(t)]$$

where r is repeatability of records, which is assumed to be .55 in USDA evaluations. Genetic, permanent environmental, and herd-sire

interaction effects of expanded records are assumed equal to those expressed in completed records. Temporary environmental effects in q (e_q) have variance

$$\text{Var}(e_q) = x^2[\text{Var}(p)] - r[\text{Var}(t)].$$

Using the identity $[\text{Corr}(p,t)]^2 = \text{Var}(p)/\text{Var}(t)$ and factoring out $\text{Var}(t)$ gives

$$\text{Var}(e_q) = \{x^2[\text{Corr}(p,t)]^2 - r\}[\text{Var}(t)].$$

The ratio $\text{Var}(e_t)/\text{Var}(e_q)$ then gives

$$w_{\text{len}} = (1 - r)/\{x^2[\text{Corr}(p,t)]^2 - r\}.$$

If x is a theoretical rather than actual expansion factor, this simplifies to

$$w_{\text{len}} = (1 - r)/(x - r).$$

Theoretical x always are ≥ 1 ; therefore, resulting w_{len} must be ≤ 1 .

Data

Data used to estimate expansion factors and lactation length weights were 48,424 first lactation records from the Pennsylvania and Wisconsin dairy records processing centers for calvings from January 1987 through July 1988. Cows included were daughters of 844 Holstein sires. Each sire was represented by six or more daughters. An additional 59 bulls provided relationship ties among the 844 and were added to the list of sires for a total of 903 bulls.

Each lactation included in the analysis had to have milk, fat, and protein information for two partial records and a 305-d record. Lactation lengths for the two records in progress (RIP) were 46 to 75 d (short) and 168 to 197 d (medium). Short and medium records were projected to a 305-d basis (7), and all records were standardized for age, calving season, and times milked daily. Estimates of genetic and error variances were obtained by REML from a multiple-trait sire model with equal design matrices. The model included effects of herd-year, genetic group, and sire (random) with sires assumed to be related (4). All nine traits (short, medium, and completed milk, fat, and protein yields) were included in one analysis.

The resulting expansion factors and length weights were applied to Ayrshire records avail-

TABLE 1. Genetic (above diagonal) and phenotypic (below diagonal) correlations between short, medium, and completed lactation records.

Trait	Lactation length	Lactation type	Lactation records		
			Short	Medium	Completed
	(d)				
Milk	46 to 75	Short	1.000	.963	.941
	168 to 197	Medium	.800	1.000	.996
	305	Completed	.795	.958	1.000
Fat	46 to 75	Short	1.000	.962	.942
	168 to 197	Medium	.741	1.000	.997
	305	Completed	.751	.944	1.000
Protein	46 to 75	Short	1.000	.947	.926
	168 to 197	Medium	.736	1.000	.993
	305	Completed	.722	.939	1.000

able for January 1990 evaluations of 123,645 cows and 1984 bulls to provide a test of the procedure. Evaluations using projected records with and without expansion were compared. A larger test using Holstein and Red and White records to investigate combined effects of expansion and adjustment for across-herd heterogeneous variance was reported by Wiggans and VanRaden (8).

RESULTS

Phenotypic and genetic correlations between short, medium, and completed records are in Table 1 for milk, fat, and protein yields. Phenotypic correlations were similar to those used to compute lactation length weights reported by Wiggans et al. (6). Theoretical expansion factors calculated with the phenotypic correlations in Table 1 for milk, fat, and protein averaged 1.76 for short records and 1.12 for medium records compared with factors of

1.72 and 1.15 from lactation length weights of Wiggans et al. (6). Genetic correlations of short with completed records averaged .936 for the three traits. Genetic correlations of medium with completed records averaged .995.

Heritabilities and sire and phenotypic standard deviations for milk, fat, and protein yields are in Table 2. As assumed in theory, heritabilities of projected records nearly equaled heritabilities of completed records divided by theoretical expansion factors. In theory, phenotypic standard deviations of projected records should equal phenotypic standard deviations of completed records multiplied by phenotypic correlation of projected with completed records. Phenotypic standard deviations of short and medium records in Table 2 were somewhat larger than expected (especially for protein) but smaller than standard deviations of completed records. Sire standard deviations of projected records averaged 70% (short) and 94% (medium) of sire standard deviations for

TABLE 2. Heritabilities and sire and phenotypic SD by lactation length.

Trait	Lactation length	Lactation type	Heritability ¹	Sire and phenotypic SD	
				Sire SD	Phenotypic SD
	(d)			(kg)	
Milk	46 to 75	Short	.11	164	983
	168 to 197	Medium	.17	231	1119
	305	Completed	.18	244	1139
Fat	46 to 75	Short	.09	5.2	34.4
	168 to 197	Medium	.13	7.4	40.4
	305	Completed	.15	8.1	41.1
Protein	46 to 75	Short	.10	4.7	29.6
	168 to 197	Medium	.13	5.7	31.2
	305	Completed	.15	6.0	31.5

¹SE = .02.

TABLE 3. Expansion factors and lactation length weights for first lactation, standard test plan records.

DIM	First lactation			Second and later lactations		
	Expansion factor	New length weight	Old length weight	Expansion factor	New length weight	Old length weight
15 to 45	1.68	.40	.47	1.90	.33	.40
46 to 75	1.43	.51	.58	1.53	.46	.53
76 to 106	1.30	.60	.67	1.34	.57	.64
107 to 136	1.20	.69	.75	1.21	.68	.74
136 to 167	1.14	.76	.81	1.13	.77	.82
168 to 197	1.09	.83	.87	1.07	.86	.89
198 to 228	1.05	.90	.92	1.04	.92	.94
229 to 258	1.03	.95	.96	1.01	.97	.98
259 to 289	1.01	.99	.99	1.01	.99	.99
290 to 305	1.00	1.00	1.00	1.00	1.00	1.00

completed records across the three traits. Genetic variance in short records as a percentage of that in completed records was highest for protein (61%) but averaged only 49%.

Heritabilities and sire standard deviations may be biased downward from selection of sires, but such selection should reduce standard deviations proportionally for each record length. Actual expansion factors for short and medium records (sire standard deviations for completed records divided by sire standard deviations for projected records) were 1.49 and 1.06 for milk yield, 1.56 and 1.09 for fat yield, and 1.28 and 1.05 for protein yield. For comparison, sire standard deviations reported by Weller (5) for annualized milk yields would give actual expansion factors of 1.45 for short records and 1.20 for medium records.

Actual amounts of expansion (expansion factor - 1) for short and medium milk, fat, and protein records averaged 61% of theoretically expected amounts of expansion based on correlations from the present study and only 53% based on correlations of Wiggans et al. (6). Actual and theoretical amounts of expansion may differ because phenotypic standard deviations of projected records differed from those expected based on phenotypic correlations. Correlations from Table 1 were available only for two lactation lengths and for first parity cows. Interpolation and extrapolation to lactation lengths, parities, and testing plans not included in the REML analysis were achieved by setting expansion factors = $1 + .60 \times$ theoretical amount of expansion (theoretical expansion factor - 1) for that record length computed from the fuller set of phenotypic correlations provided by Wiggans et al. (6).

The resulting expansion factors, previous lactation length weights used by Ustandard deviations, and new lactation length weights are in Table 3. New lactation length weights were calculated from the expansion factors in Table 3 using the formula $(1 - r)/(x - r)$. For first lactations, lactation length weights of .51 for short records and .83 for medium records resulted from expansion factors of 1.43 and 1.09. Current evaluation programs apply the same lactation length weights and expansion factors to milk, fat, and protein, but separate adjustments could be justified.

Expanded records were calculated by subtracting management group mean from projected yield, multiplying the remainder by the appropriate expansion factor, and adding back the management group mean. For example, a first lactation cow with an RIP at 50 DIM might have a projected record (standardized to a twice daily milking, mature equivalent, 305-d basis) of 10,500 kg for milk yield. Mean milk yield of her management group was 8000 kg. Her expanded record would be

$$8000 + 1.43(10,500 - 8000) = 11,575 \text{ kg}$$

TABLE 4. Change in PTA milk and reliability (REL) for 1276 Ayrshire cows born in 1987.

	PTA Milk		Mean REL
	\bar{X}	SD	
	— (kg) —		(%)
Projected	48	153	37.3
Expanded	50	167	36.7
Difference	2	19	-.6

TABLE 5. Change in PTA milk and reliability (REL) for 43 Ayrshire bulls with 30% of daughters or more with first lactation records in progress.

	PTA Milk		Mean
	\bar{X}	SD	REL
	— (kg) —		(%)
Projected	156	217	58.6
Expanded	164	233	58.2
Difference	8	21	-.4

where 1.43 is the expansion factor from Table 3. For inclusion in the animal model, this expanded record receives a weight of .51 compared with the weight of .58 previously applied to the projected yield.

Changes in PTA milk and reliability from applying expansion factors and lactation length weights to Ayrshire data are in Tables 4 and 5. Table 4 documents changes for the 1276 cows born in 1987, most of which had PTA computed from first lactation RIP. Table 5 shows changes for the 43 bulls with at least 30% of daughters with first lactation RIP. Use of expanded records increased standard deviations of PTA milk by 14 kg for the recent Ayrshire cows and by 16 kg for the bulls with many new daughters. Reliability decreased by an average of about .5 because previous reliability for animals evaluated primarily from RIP were overestimated.

The PTA calculated from projected and expanded records were correlated by .997 for cows and .998 for bulls. If all differences between new and old procedures represent improved accuracy, a .2 to .3% increase in annual genetic progress is expected using expanded rather than projected records.

CONCLUSIONS

Projected and completed records have unequal genetic variances, but variances can be standardized by simple expansion of projected records. Expansion factors calculated as reciprocals of squared phenotypic correlations were larger than those calculated directly as ratios of genetic standard deviations. Reciprocals of squared phenotypic correlations were

adjusted downward to match ratios of genetic standard deviations to provide expansion factors for record types not included in the genetic parameter estimation. High genetic correlation of short and completed yields and equal genetic variance provided by expanded records make even extremely short RIP easy to model in genetic evaluations. Increased variance of expanded RIP is offset partially by reduced weight for these records in evaluations. Use of expanded records has little effect on most animals' evaluations but has larger benefits for younger animals newly available for selection. The expanded records and reduced lactation length weights should lead to more stable and accurate evaluations, particularly for animals evaluated from first lactation RIP. These expansion factors and length weights were implemented for all breeds beginning with January 1991 USDA genetic evaluations. Mexican evaluations also have included expanded records since 1990 (3).

ACKNOWLEDGMENTS

The National Cooperative DHI Program supplied the data for this research.

REFERENCES

- 1 Danell, B. 1990. Genetic aspects of different parts of lactation. Proc. 4th World Congr. Genet. Appl. Livest. Prod., Edinburgh, XIV:114.
- 2 Dickinson, F. N., H. D. Norman, R. L. Powell, L. G. Waite, and B. T. McDaniel. 1976. Procedures used to calculate the USDA-DHIA Modified Contemporary Comparison. Page 18 in USDA Prod. Res. Rep. No. 165, Washington, DC.
- 3 Powell, R. L., and G. R. Wiggans. 1991. Animal model evaluations for Mexican Holsteins. *J. Dairy Sci.* 74:1420.
- 4 VanRaden, P. M. 1986. Computational strategies for estimation of variance components. Ph.D. Diss., Iowa State Univ., Ames.
- 5 Weller, J. I. 1988. Inclusion of partial lactations in the genetic analysis of yield traits by differential weighting of records. *J. Dairy Sci.* 71:1873.
- 6 Wiggans, G. R., I. Misztal, and L. D. Van Vleck. 1988. Implementation of an animal model for genetic evaluation of dairy cattle in the United States. *J. Dairy Sci.* 71(Suppl. 2):54.
- 7 Wiggans, G. R., and R. L. Powell. 1980. Projection factors for milk and fat lactation records. *USDA DHI Lett.* 56(1), Beltsville, MD.
- 8 Wiggans, G. R., and P. M. VanRaden. 1991. Method and effect of adjustment for heterogeneous variance. *J. Dairy Sci.* 74:4350.
- 9 Wiggans, G. R., and L. D. Van Vleck. 1979. Extending partial lactation milk and fat records with a function of last-sample production. *J. Dairy Sci.* 62:316.