GENETICS AND BREEDING

Comparison of Test Interval and Best Prediction Methods for Estimation of Lactation Yield from Monthly, a.m.–p.m., and Trimonthly Testing

H. D. NORMAN,*,[1] P. M. VanRADEN,* J. R. WRIGHT,* and J. S. CLAY[†]
*Animal Improvement Programs Laboratory, Agricultural Research Service, USDA, Beltsville, MD 20705-2350
[†]Dairy Records Management Systems, Raleigh, NC 27695

ABSTRACT

A method with best prediction properties that condenses information from all test days into measures of lactation yield and persistency has been proposed as a possible replacement for the test interval method and projection factors. The proposed method uses previously established correlations between individual test days and includes inversion of a matrix for each lactation. Milk weights that were representative of monthly, a.m.–p.m., and trimonthly test plans were examined to compare the accuracy of best prediction and test interval methods for estimating lactation yield. Individual milk weights or daily yields of 658 Canadian cows in 17 herds were selected to correspond to test intervals for 100,000 US cows. For a.m.–p.m. testing, the initial milk weight that was credited was selected randomly from the a.m. or p.m. milking and was alternated thereafter. Trimonthly credits were from one of the first three designated test day weights, selected randomly, and each third designated test weight thereafter. Correlations between 305-d actual lactation yield and lactation estimates by the test interval method were 0.97, 0.96, and 0.93 for monthly, a.m.–p.m., and trimonthly testing, respectively. Corresponding correlations for the best prediction method were 0.97, 0.97, and 0.93. Standard deviations of differences between estimated and 305-d actual yields for monthly, a.m.–p.m., and trimonthly testing were 373, 400, and 546 kg, respectively, for best prediction regressed on herd mean, which was a reduction in estimation error of 4, 6, and 10% over the test interval method. The advantage of best prediction was moderate if two milk weights were recorded monthly and was larger if testing was less frequent. Advantages also were found for fat and protein yields estimated by multitrait best prediction for records with reduced component sampling.

(**Key words**: test interval method, best prediction, lactation yield)

Abbreviation key: **AP** = a.m.–p.m., **BP** = best prediction, **MN** = monthly, **TIM** = test interval method, **TRI** = trimonthly.

INTRODUCTION

Historically, most cows enrolled in a milk recording program in the US have had milk weights recorded monthly (14). However, practices for collecting milk and component data (fat and protein) have changed rapidly. More than 30 innovative test plans were proposed and introduced between 1989 and 1995 (P. Dukas, 1995, personal communication). Many of the new plans require less labor for recording milk weights and collecting component samples (fat and protein), thereby reducing cost to the producers. These plans differ widely in the number of milk weights recorded and the number of component samples taken. Often cows in innovative testing programs are tested less frequently than monthly, and, in some cases, herds have as few as four milk component samples tested per year. Technological development has made electronic recording of each milking feasible; therefore, some innovative test plans have milk weights recorded more than once per month.

In 1969, the test interval method (**TIM**) (8) replaced the centering date method in the US for estimating 305-d lactation milk yield because TIM produced more accurate estimates than the centering date method when milk weights and component samples were obtained each month (5). With the centering date method, a cow was credited for the first test period with the yield on first test day multiplied by the number of days since calving. The yield credit for each successive test period was the yield produced on the successive test day multiplied by the number of days since the previous test. The credit for the last test period for lactations with <305 DIM was the yield

Received May 14, 1998.
Accepted November 23, 1998.
[1]Reprint requests.

on last test day multiplied by the number of days from the last test day through the end of lactation. For lactations with ≥305 DIM but with no test day at ≥305 DIM, the credit for the last test period was the yield on the last test day multiplied by the number of days from the last test day through d 305. The centering date method generally overestimated actual yields until peak lactation as well as yield during the period following the last test day but underestimated yields for other test periods.

The test interval for TIM is the period immediately after a test day through the following test day and is composed of two equal parts (2). Production credits for the first half of the test period are assumed to be the same as the earlier test day yield, and credits for the last half are assumed to be the same as the later test day yield. For first and last test intervals, yield credits are calculated the same as for the centering date method. During the 1970s, TIM was improved by adjusting credits for the first and second test intervals for the nonlinear shape of the lactation curve and by adjusting the last test for a continuation of the expected decline (11). Those adjustments eliminated biases from overestimation of the credits for first and last test intervals and from an underestimation of the credits for the second test interval.

In 1985, Norman et al. (6) showed that extending lactation yields to 305 d for those cows that remained in the herd but had discontinued milking before 305 d produced higher heritabilities and repeatabilities for lactation yield than if those records had not been extended. This finding was the basis for standardizing all lactation records to a length of 305 d for use in calculating US genetic evaluations. However, breed associations continue to show actual yields on pedigrees as the expression of phenotype. In several other countries, records terminated at <305 d are either not used or are not extended in their genetic evaluation methods (4). Henderson et al. (3) showed that the failure to include all records on which selection is based produces biased genetic evaluations.

Recently, VanRaden (12) developed a procedure to calculate the accuracy of records from the current (or any proposed) test plans by considering the number and distribution of tests. The method also provided an estimate of lactation milk and component yields from test day data. The procedure has best prediction (**BP**) properties and condenses information from individual test days into lactation measures of yield and persistency. The BP method uses previously established correlations between individual test days (7) and includes inversion of a matrix for each lactation record. The method can provide a prediction of the yield for any day of the lactation or the lactation total for any length through 305 d. If the mean yield for the herd is available, that yield can be used to improve the accuracy of the prediction. Otherwise, the prediction assumes the cow is from an average herd and includes mean breed yield. A theoretical basis exists for hypothesizing that a BP method is more accurate than TIM. The objective of this study was to test this hypothesis using daily yield data. Differences in the accuracy of prediction also were exam-ined for multitrait and single-trait analyses, which is pertinent to test plans in which components are recorded less frequently than milk weights.

MATERIALS AND METHODS

Estimates from BP methods and TIM were compared with actual milk yields for a wide range of milk recording frequencies to determine which method was more accurate for estimating lactation yield. Measures of accuracy were correlations between actual and estimated yields and standard deviations of differences between actual and estimated yields. Of particular interest was the determination of the accuracy of the methods for test plans in which milk weights or component samples were not collected on each milking on the test day or were recorded less frequently than once per month. This study included testing environments for cows on 1) traditional monthly (**MN**) testing (i.e., herds for which milk weights are taken approximately 1 d/mo), 2) a.m.–p.m. (**AP**) testing (i.e., herds for which only one of the two milk weights is recorded on the monthly test date), and 3) trimonthly (**TRI**) testing (i.e., herds for which weights are taken 1 d every 3 mo). The estimates from the TIM and BP methods were compared with two measures of actual milk yield within each testing environment.

Individual daily milk weights were available for 658 Canadian Holstein cows in 17 herds. For all cows, daily yields had been collected since the beginning of lactation, and lactation lengths were ≥250 d. Individual milk weights had been recorded for ≥90% of milkings during each lactation.

One measure of actual lactation yield (≤305 d) was based only on reported milk weights; the maximum DIM was 305. Some milk weights that were not reported in the data were estimated before actual lactation yields were derived. If only one of the two daily milking weights was available, daily yield was calculated as twice the single-milking weight. Milking times were not recorded in the Canadian data set; therefore, adjustments for milking time, as suggested by Shook et al. (10), were not possible. If both daily milkings were missing, daily yields were estimated using linear interpolation if yields were reported for the days immediately before and after missing test days. If daily yield was missing for the first or last day of lactation, missing daily yield was considered to be the same as the yield for the closest day. Using the actual recorded and estimated milk weights, when individual milkings were not reported, ≤305-d actual lactation yields were derived by summing daily yields.

A second measure of actual lactation yield (305 d) was based on reported milk weights plus a credit for production through 305 d for cows with lactation lengths of <305 d. This credit is what the cow could have produced if management had not terminated the lactation. The 305-d actual lactation yields were derived by summing daily yields and projected credits.

All yield information came from the 658 Canadian cows; however, all testing intervals came from 100,000 US cows enrolled in MN test plans (1). Test intervals were determined from the DIM reported for test days of the US cows. Individual milk weights or daily yields of the Canadian cows were selected to correspond with test intervals of the US cows. The Canadian data file was used multiple times to construct 100,000 observations. Although individual daily yields were used several times (approximately five) in constructing this number of observations, only a small percentage of the possible daily yield combinations were expected to be repeated, as most would not have been expected to be selected at all. This replication was useful in extracting more information from the data set and had no apparent disadvantage. If significance testing had been done, accounting for the replication would have been necessary. Lactation yields were estimated using both TIM and a BP method to compare the accuracy of estimation. Herd means that were calculated for the 17 Canadian herds enabled two measures of BP to be calculated for each record: one using mean breed yield in the regression and the other using mean herd yield.

For AP testing, an initial milk weight on the first test day for the lactation was selected randomly from the a.m. or p.m. milking, and alternate milkings were used on subsequent test days. For TRI testing, the first daily yield that was used was selected randomly from the first three designated test days, and subsequent daily yields came from each third designated test day.

A major disadvantage of TIM is that observations for correlated traits cannot be used to predict missing values. Multitrait BP allows missing component yields to be predicted from milk yields. Because daily yields of fat and protein were not available, TIM and BP methods could not be compared with true lactation yields of fat and protein. However, single-trait and multitrait BP methods were compared theoretically to quantify the value of multitrait procedures when some traits were missing.

Expected correlations were obtained for both single-trait and multitrait BP methods by pre- and postmultiplying the phenotypic variance inverse by the covariance of lactation and test day yields. Fat and protein yields were assumed to be as correlated as milk yields that were separated by the same number of days. On the same day, yields of milk and fat were assumed to be correlated by 0.65, milk and protein by 0.88, and fat and protein by 0.72. Those correlations were derived from monthly observations of milk, fat, and protein yields from the same Canadian cows. The multitrait expected correlations of true lactation yield with estimated lactation yield are used as a data collection rating by the US industry after the values are squared and adjusted so that monthly testing plans are rated 100% (12).

RESULTS AND DISCUSSION

Means and standard deviations for actual and estimated lactation milk yields in MN, AP, and TRI testing are presented in Table 1. As expected, means were higher (362 to 447 kg for MN testing, 264 to 331 kg for AP testing, and 303 to 361 kg for TRI testing) for all procedures that credited yield through 305 d; AP and TRI means of actual and estimated yields were similar to MN means. No obvious explanation existed for the small differences between testing environments.

Standard deviations (Table 1) were lowest for the two BP estimates for all testing environments because those methods regressed outlier test day yields either toward breed mean or herd mean. Each corresponding standard deviation was greater for AP than for MN testing except for the BP method considering herd mean. The large variation is not unexpected considering that no milking times were recorded in the Canadian data set and that daily yield for AP testing was assumed to be twice the single-milking weight. For TRI testing, the standard deviation for TIM was considerably larger than those from MN and AP testing; however, the standard deviations for both BP methods were smaller than that for MN testing. The standard deviation of estimated yield from TRI testing was smaller than that from AP

TABLE 1. Actual and estimated lactation milk yields for traditional monthly (MN), a.m.–p.m. (AP), and trimonthly (TRI) testing.

Yield	MN Testing		AP Testing		TRI Testing	
	(kg)					
	$\overline{X}$	SD	$\overline{X}$	SD	$\overline{X}$	SD
Actual						
≤305 d	8302	1554	8421	1579	8299	1553
305 d	8664	1538	8716	1543	8660	1535
Estimated						
Test interval	8697	1545	8715	1564	8669	1622
Best prediction						
Breed mean	8748	1500	8752	1512	8602	1453
Herd mean	8749	1507	8685	1461	8604	1473

testing for BP regressed to breed mean but was similar for BP when herd mean was considered.

The mean for test interval was 32.8 d for MN testing (SD = 8.8 d). Because the standard deviation for test interval was higher than expected, the Dairy Records Management System (Raleigh, NC) was contacted to determine the mean and standard deviation for test intervals from field data. Of the 14,622 herds with records processed by Dairy Records Management System, the mean for test interval was 33.4 d (SD = 9.8 d). Those herds included about 100 herds using test plans that did not schedule tests in consecutive months.

Correlations between actual and estimated milk yields for observations are presented in Table 2 for MN, AP, and TRI testing. For all types of testing, the correlation between ≤305-d and 305-d actual yields was 0.98. For MN testing, correlations between actual and estimated yields were high and the same (to two decimal places) for both estimation methods (0.96 with ≤305-d actual yield and 0.97 with 305-d actual yield). The corresponding correlations for AP testing

TABLE 2. Correlations among actual and estimated lactation milk yields in traditional monthly (MN), a.m.–p.m. (AP), and trimonthly (TRI) testing.

Yield	Actual yield		Estimated yield	
	≤305 d	305 d	Test interval method	Best prediction method
	MN			
Actual				
≤305 d	1.00	0.98	0.96	0.96
305 d		1.00	0.97	0.97
Estimated				
Test interval method			1.00	1.00
Best prediction method				1.00
	AP			
Actual				
≤305 d	1.00	0.98	0.95	0.96
305 d		1.00	0.96	0.97
Estimated				
Test interval			1.00	1.00
Best prediction				1.00
	TRI			
Actual				
≤305 d	1.00	0.98	0.92	0.93
305 d		1.00	0.93	0.93
Estimated				
Test interval			1.00	0.99
Best prediction				1.00

were the same as for MN testing except that the correlations between TIM and actual yields were lower by 0.01. For TRI testing, correlations between actual and estimated yields were lower (0.92 to 0.93) than for MN and AP testing. This result is in agreement with previous reports by Meinert (1995, unpublished National DHIA minutes) and McDaniel (5). Obviously, as the number of recorded daily weights decreases, the ability of the recorded information to predict actual lactation yield accurately also decreases.

Correlations of BP estimates with other variables were the same for BP regressed to either breed or herd mean regardless of testing environment; therefore, the correlations shown in Table 2 are appropriate for either BP method. Because yields came from only 17 Canadian herds, the differences in accuracy between BP with herd mean versus breed mean might vary for data sets based on national populations, depending on whether the 17 herds studied had more or less variation in yield than the entire population.

The correlations between TIM and BP estimated yield (Table 2) were extremely high (≥0.99) for all testing environments, which suggested that predictions of 305-d yield from the two methods would be similar. Correlations to three decimals (not shown in Table 2) between TIM and BP estimates considering breed or herd mean were 0.998 for both BP estimates for MN testing, 0.995 for herd mean BP and 0.998 for breed mean BP for AP testing, and 0.994 for both BP estimates for TRI testing. The correlations were high partly because of the large variation among yields of individual cows. Therefore, the differences between TIM and BP estimates were considered to be important despite the high correlations.

Standard deviations of differences between estimated and actual yields are shown in Table 3. The standard deviations of differences between either BP method and both measures of actual yield were smaller than for differences between TIM and actual yield for all testing environments. For comparisons of estimated yield with 305-d actual yield, standard deviations of differences for both BP methods were 4% smaller than for TIM for MN testing, 6% smaller for AP testing, and 10% smaller for TRI testing, which indicates that both BP methods have a moderate advantage over TIM despite correlations between TIM and BP estimates of nearly 1.0. The advantage increased for testing environments in which weights and samples were recorded at less optimal intervals and frequencies. The main advantage of a BP method over TIM is that lactation records are regressed toward a population mean when few weights or samples are available. A regressed TIM might rival a BP method. However, TIM cannot be used to predict missing values for correlated traits, and a method to regress the lactation yield is not obvious, particularly when the testing intervals are not equally spaced throughout the lactation.

TABLE 3. Standard deviation of differences between actual and estimated lactation milk yields in traditional monthly (MN), a.m.–p.m. (AP), and trimonthly (TRI) testing.

	SD of Difference from actual yield		
		Best prediction method	
Testing environment and actual yield	Test interval method	Breed mean	Herd mean
	(kg)		
MN			
≤305 d	451	438	437
305 d	388	374	373
AP			
≤305 d	483	461	463
305 d	425	398	400
TRI			
≤305 d	635	582	579
305 d	606	549	546

The opportunity to increase the accuracy of predicting 305-d yield by 4 to 10% is important and could result in large economic returns in the efficiency of milk yield. Often, culling decisions are based on a cow's actual yield, real producing ability, or predicted producing ability (13). Because the repeatabilities of milk and component yields are near 50%, greater accuracy in calculating those rankings for early parities would mean that cows selected for survival by herd managers would yield more during later lactations than if the rankings had been based on less accurate TIM estimates.

Advantages of multitrait over single-trait BP prediction of lactation yield are illustrated in Table 4. Multitrait correlations of predicted and true yield were similar to those reported by Schaeffer and Jamrozik (9) using a random regression approach. The single-trait correlations in Table 4 were similar for fat and protein yields but were lower than for milk yield because component samples were less numerous than milk weights. Multitrait correlations were higher for protein yield than for fat yield because milk yield is more highly correlated with protein yield than with fat yield.

For test plans with components recorded, multitrait correlations were from 1 to 5% higher than single-trait correlations for fat yield and from 2 to 9%

TABLE 4. Expected correlations of actual with predicted lactation yields from single-trait and multitrait best prediction (BP) methods for test plans with reduced component sampling.

Recording frequency		Single-trait BP			Multitrait BP		
Milk	Component	Milk	Fat	Protein	Milk	Fat	Protein
Monthly	None	0.98	0.00	0.00	0.97	0.65	0.88
	Trimonthly	0.98	0.89	0.89	0.98	0.92	0.94
	Bimonthly	0.98	0.92	0.92	0.98	0.94	0.95
5-d LER[1]	Bimonthly	1.00	0.92	0.93	1.00	0.94	0.95
150-d RIP[2]	Trimonthly	0.93	0.79	0.81	0.93	0.83	0.88
	Bimonthly	0.93	0.85	0.87	0.93	0.87	0.89

[1]Labor efficient records in which daily weights are averaged over 5 d.

[2]Records in progress.

higher for protein yield. With no component testing, the single-trait correlations of 0 were far below the multitrait correlations of 0.65 for fat yield and 0.88 for protein yield. Those latter correlations equal the assumed phenotypic correlation of each trait with milk yield. For milk yield, multitrait correlations were not higher than single-trait correlations because milk weights were always recorded when components were sampled.

CONCLUSIONS

Currently TIM is the accepted procedure in the US for estimating lactation yield from test day data. Daily yields from 658 Canadian Holsteins and information on US testing frequencies from 100,000 cows were used to determine whether BP methods could estimate lactation yields more accurately than TIM. For the traditional environment of MN tests, a moderate decrease in estimation error (4%) was found for BP methods compared with TIM in estimating 305-d actual milk yield. For AP and TRI testing, the decreases in estimation error were 6 and 10%, respectively. Because 60% of the US cows on test are enrolled in an AP plan in which intervals between tests have increased and components are not sampled at each test, new procedures that more accurately predict lactation yields for such plans (such as the BP method) should be considered as replacements for TIM. However, further research is needed to examine the accuracy of BP for projecting records in progress to 305 d.

A theoretically more accurate BP procedure was shown also to be more accurate with empirical data. The use of multitrait BP increased the accuracy of estimated fat and protein yields by 1 to 9% when component samples were obtained less frequently than milk weights. Although the results are promising, further research is needed to determine the appropriateness of multitrait BP for US milk, fat, and protein yields and SCS based on a larger set of daily data from US cows or from totally independent herds. Dairy record processing centers should evaluate the merit of implementing an alternative method to TIM to make better use of data from new milk recording plans.

ACKNOWLEDGMENTS

The authors thank R. K. Moore, Québec Dairy Herd Analysis Service (Ste. Anne de Bellevue, PQ, Canada); L. R. Schaeffer, University of Guelph (Guelph, ON, Canada); and T. R. Meinert, formerly of National DHIA (Columbus, OH), for file preparation and transfer of Canadian data to USDA (Beltsville, MD). Estimates of lactation yield by TIM were calculated by Dairy Records Management System (Raleigh, NC). The assistance of K. R. Butcher, North Carolina State University (Raleigh); R. H. Miller, University of Maryland (College Park); and P. D. Miller, National DHIA, for manuscript review also is appreciated.

REFERENCES

1 Benson, R. H. 1985. The NCDHIP record plans. Natl. Coop. DHI Prog. Handbook, Fact Sheet A-4. Ext. Serv., USDA, Washington, DC.
2 Empet, N. B. 1985. Official Dairy Herd Improvement rules. Natl. Coop. DHI Prog. Handbook, Fact Sheet E-1. Ext. Serv., USDA, Washington, DC.
3 Henderson, C. R., O. Kempthorne, S. R. Searle, and C. M. von Krosigk. 1959. The estimation of environmental and genetic trends from records subject to culling. Biometrics 15:192–218.
4 International Bull Evaluation Service. 1988. Sire evaluation procedures for dairy production traits practised in various countries, 1988. Bull. No. 3. Dep. Anim. Breeding Genet., SLU, Uppsala, Sweden.
5 McDaniel, B. T. 1969. Accuracy of sampling procedures for estimating lactation yields: a review. J. Dairy Sci. 52: 1742–1761.
6 Norman, H. D., F. N. Dickinson, and J. R. Wright. 1985. Merit of extending completed records of less than 305 days. J. Dairy Sci. 68:2646–2654.
7 Norman, H. D., P. M. VanRaden, J. R. Wright, and T. R. Meinert. 1996. Mathematical representation of relationships

between daily milk yields. J. Dairy Sci. 79(Suppl. 1): 143.(Abstr.)
8 Sargent, F. D., V. H. Lytton, and O. G. Wall, Jr. 1968. Test interval method of calculating Dairy Herd Improvement Association records. J. Dairy Sci. 51:170–179.
9 Schaeffer, L. R., and J. Jamrozik. 1996. Multiple-trait prediction of lactation yields for dairy cows. J. Dairy Sci. 79: 2044–2055.
10 Shook, G. E., E. L. Jensen, and F. N. Dickinson. 1980. Factors for estimating sample-day yield in am–pm plans. DHI Lett. 56(4):25–30.
11 Shook, G. E., L. P. Johnson, and F. N. Dickinson. 1980. Factors for improving accuracy of estimates of test-interval yield. DHI Lett. 56(4):9–24.
12 VanRaden, P. M. 1997. Lactation yields and accuracies computed from test day yields and (co)variances by best prediction. J. Dairy Sci. 80:3015–3022.
13 VanRaden, P. M., and G. R. Wiggans. 1991. Derivation, calculation, and use of national animal model information. J. Dairy Sci. 74:2737–2746.
14 Voelker, D. E. 1981. Dairy Herd Improvement Associations. J. Dairy Sci. 64:1269–1277.