

INTERBULL OPEN MEETING, PRAGUE 1995

Genetic Evaluation with a Multiple Lactation Test Day Model for SCS and Production Traits

R. REENTS*, J. C. M. DEKKERS, and L. R. SCHAEFFER

Centre for Genetic Improvement of Livestock, University of Guelph

* present address: VIT, Heideweg 1, 27283 Verden Germany, Email: rreents@rzv-srv3.vit.de

ABSTRACT

A multiple lactation test day model was applied to predict genetic merit for somatic cell score and protein yield in Canadian Holstein cattle. The model for test day genetic evaluation included a fixed herd-test-date effect, fixed regressions on functions of days in milk, random permanent environmental effects within lactation, random animal genetic effects, and residual effects. Records from the first three lactations were used and treated as different traits. Procedures for this model were developed for national genetic evaluation for somatic cell score and were found to be practical even for a much larger dataset from western Germany. Use of starting values from the previous genetic evaluation run reduced the number of rounds necessary to reach convergence. Test day models were compared to several single trait models based on lactation averages in terms of ranking of animals. Differences between EBV from the test-day model and EBV from a lactation average repeatability model were small for bulls with many daughters, but differences with EBV from a single trait first lactation average model were large. Differences were smaller for SCS than for protein yield.

OBJECTIVES

The objectives of this study were:

- a) to develop procedures for national genetic evaluation based on multiple lactation test day models and to apply them to field data (from Canada and Germany)
- b) to compare test-day models to lactation average models for SCC and for protein yield.

Two different test day models were compared to models for genetic evaluation based on lactation measures using the data set from Canada (data set 1). Results on computational aspects will be presented for the larger German data set (data set 2).

MATERIALS AND METHODS

Data Data consisted of test day records for SCC and production traits from data records processing centres in Canada and Germany, which were analyzed separately. For Canada, data included records from test days from 1987 until fall of 1994. For Germany, the time span was from 1990 to summer 1995. For each test day, SCC was transformed to SCS to achieve normality and homogeneity of variances.

Table 1 shows numbers of records and numbers of herd-test-date levels and cows. Following edits, 5,505,672 records on 392,487 cows remained for Canada and 22,617,153 records on 1,487,279 cows remained in the German data set. Pedigree was completed for cows with identification of dam and maternal grandsire from national pedigree files. Pedigrees for bulls with daughter records or granddaughter records were completed for several generations. Unknown parents were assigned to phantom parent groups, grouped by birth year of offspring (5 years per interval).

Models

For genetic evaluation of test-day observations, a multiple trait test day model with repeated

observations within each lactation was used. Table 2 displays (co)variance matrices of additive genetic (G_0), permanent environmental (PE_0), and residual (R_0) effects, which were estimated using Gibbs sampling procedures on subsets of the analysed data sets. Material and methods for the Canadian dataset are described in a paper by Reents et al. (1995a). Variance components for Germany were estimated applying the same methodology using a dataset of 26,216 German Holstein cows with 362,478 test day records from lactations 1 to 3.

Two models were applied for analysis of test-day records. The first model (TD1) was:

$$y_{ijkmn} = \text{HTD}_{im} + A_{jm} + P_{jm} + AS_{km} + b_{km1}(D/c) + b_{km2}(D/c)^2 + b_{km3}\ln(c/D) + b_{km4}[\ln(c/D)]^2 + e_{ijkmn}$$

where y_{ijkmn} is the n^{th} test day observation of the j^{th} cow in parity m ; HTD_{im} is a fixed herd-test-date effect; A_{jm} is an animal additive genetic effect (random); P_{jm} is a within lactation permanent environmental effect to account for common environmental effects associated with all test-day records of the j^{th} cow in lactation m (random); AS_{km} is an age-season subclass mean effect in parity m ; b_{km1} and b_{km2} are regression coefficients on the linear and quadratic effects of D/c , where D is days in milk and $c=381$; b_{km3} and b_{km4} are regression coefficients on the linear and quadratic effects of $\ln(c/D)$; e_{ijkmn} is a random residual effect. Regression coefficients were estimated within 24 parity*age of calving*season groups.

Test-day model TD2 was very similar to model TD1, except contemporary groups for second and third lactation records from a specific herd-test-date were combined into a common herd-test-day class to increase the size of subcells.

Three genetic evaluation models based on lactation average SCS or 305-d protein yield were used for comparison. Model REP1 was a single trait repeatability model:

$$y_{ijkl} = \text{HYS}_i + \text{PA}_j + P_k + A_k + e_{ijkl}$$

where y_{ijkl} is the l^{th} lactation average of the k^{th} cow, HYS_i is a fixed herd-year-season effect, PA_j is a fixed parity by age effect, P_k is a random permanent environmental effect across lactations of cow k , A_k is a random additive genetic effect of animal k , and e_{ijkl} is a random residual effect. Residual variances were scaled according to the number of samples included in the lactation average, based on appropriate weighting factors. Only averages based on at least two test day records were included. To compute lactation average SCS, test day records were adjusted for stage of lactation by additive pre-adjustment factors. Variance components used were based on a heritability of .11 and a repeatability of .27 for SCS and a heritability of .33 and a repeatability of .5 for protein yield, as currently used in routine evaluation in Canada.

Model REP2 was similar to REP1, but only lactation yields based on at least five test-day observations were used. All observations received the same weight in the evaluation. Variance ratios were the same as used in model REP1.

Model ST was similar to REP2, except that only first lactation data were included. Only lactation yields based on at least five test-day results were used. Pre-adjustment for stage of lactation was done with the same factors as for models REP1 and REP2. A heritability of .11 for SCS and of .33 for protein yield was used.

Models TD1 and TD2 provide separate EBV for SCS in the first three lactations. EBV can be combined into an overall EBV for SCS based on their economic values, with index weights derived based on Schneeberger et al. (1991) as: $b = G_0^{-1} G_T v$, where G_T is a matrix of genetic (co)variances between SCS in the first 3 lactations and traits in the breeding goal and v is a vector of economic values of traits in the breeding goal. The breeding goal and economic values used were as developed by Kolstad and Dekkers (1994, Unpublished) and included four traits: clinical mastitis and subclinical mastitis (SCS) in first and second lactations. Economic values for second lactation traits accounted for the expression of traits in all later lactations, assuming a genetic correlation of one. Resulting index weights, standardized to sum to one, were .26, .66, and .08 for EBV for SCS in lactations 1, 2, and 3. For protein yield equal weights were put on each lactation EBV.

Computing strategies used for iterative solution of large scale test day animal models were described in detail by Reents et al. (1995b). The important key for processing a large number of test day records is the use of efficient input and output routines of C-HP (fread) in fortran programs. This allowed solving for each effect in the model by iteration on data (Schaeffer and Kennedy, 1986). Considerable amount of random access memory was saved by using an implicit representation of the mixed model equations for a multiple trait animal model, as described by Tier and Graser (1991). In this method unique diagonal blocks of random effects are stored in memory only once and addressed by pointers when processing the permanent environment or animal effect. Animal effects were solved via Second Order Jacobi iteration with a relaxation factor of .7. All other effects were solved by the Gauss-Seidel method. The relative difference between consecutive solutions prior to relaxation was used to monitor convergence.

RESULTS AND DISCUSSION

Memory requirements were found to be no limitation for application of a multiple lactation test-day model to the national datasets of Canada and Germany. Although model TD2 for the German data set consisted of 14,704,144 equations, and 22 mio test day records were processed three times in every round of iteration, less than 200 MBytes of memory were required. Due to the implicit representation of the mixed model equations, test-day models TD1 and TD2 required only slightly more memory than lactation average models. One round of iteration took about 12 minutes CPU time on a HP 9000-891 workstation for model TD2 for the German data set. Processing time decreased by 25% if all test-day-records were stored in memory. Each record required 18 bytes of memory, thus overall demand of random access memory increased by about 400 MB for dataset 2. In this data set, 3104 different diagonal blocks for permanent environment effects on 1.4 mio cows were found. Consideration of inbreeding for construction of the A-1 increased number of different elements in A-1, but still only a total of 75005 different diagonal blocks for animals were found, although the complete dataset comprised of nearly 2.4 mio animals.

From detailed studies on dataset 1 (Table 3) it was concluded that 250 to 300 rounds of iteration were sufficient to reach stable solutions (relative difference, expressed on a log10 scale, reached a value of -4.0). A similar level of convergence was also reached for dataset 2 after 250 rounds.

The same mixed model equations (dataset 1) were also solved using solutions from a simulated previous (6 mo earlier) genetic evaluation run as starting values. The same level of convergence, which was reached after 300 rounds without starting values, was now reached after 150 rounds. Benefits from the use of starting values will likely be larger when more frequent evaluations (and therefore less new observations) are computed.

In model TD1, records from different lactations on the same sample day were fitted in different herd-test-date groups. For third lactation herd-test-date levels, subcell size becomes small in small herds and, therefore, the effective contribution of these records to genetic evaluation is small when herd-test-date effects are considered as fixed. For dataset one, 14.3% of all third lactation records were grouped in herd-test-date levels of size 1 or 2. Subcell sizes for herd-test-date levels for lactation 2 were larger, but still 6.1% of all second lactation records were grouped in herd-test-date levels of size 1 or 2. Assigning records from second and third lactation to the same herd-test-date group, as in model TD2, resolved this problem, giving only 2.9% (3.4% for the German dataset) of second and third lactation records in herd-test-date levels of size 1 or 2. Differences between EBV from models TD1 and TD2 were extremely small for the combined EBV over all three lactations and therefore in further tables only results from TD2 will be used.

Table 4 shows correlations of combined EBV for SCS and protein yield from model TD2 with EBV from the three lactation average models for Canadian data. For SCS, differences between EBV for bulls from test-day models and lactation average models diminished with increasing numbers of daughters (Table 4). Lowest correlations were found with the first lactation model ST;

even with more than 100 daughters, correlations were as low as .86. Repeatability models REP1 and REP2 showed better agreement but correlations for bulls with 26 to 50 daughters, which might be characterized as young test bulls with their first crop of daughters, were .95. Correlations between cow EBV were in a range similar to correlations for bulls with a low number of daughters. For protein yield correlations were in general smaller than for SCS from this study and also lower than reported by Swalve (1995) for a study with 1st lactation German production data. It has to be investigated if differences in the present study are due to the difference in heritabilities used in the test day models (.25, .23, and .20 for lactations 1 to 3) compared to a heritability of .33 for lactation protein yield or if in general larger differences between test day models and lactation models can be found for multiple lactation TD models applied to production traits compared to SCS.

For production traits, most attention is paid to the ranking of animals with the most desirable (= highest) EBV. For SCS, focus would be mostly on avoiding bulls with high EBV for SCS, which would be undesirable. Table 5 displays the agreement of the ranking of bulls with highest EBV for SCS and protein yield from test day model TD2 to the ranking based on EBV from lactation average models REP1, REP2, and ST. As already indicated by the correlation between EBV, differences in ranking were largest for bulls with a low number of daughters. Agreement of TD2 was better with the repeatability models REP1 and REP2 than with model ST. With model ST, only 5 to 12 out of the 20 highest ranking bulls (for SCS) and 6 to 8 out of the 20 best bulls for protein yield from TD2 were ranked in the same group. In general differences between TD models and lactation average models were larger for protein yield than for SCS.

REFERENCES

- Reents, R., J. Jamrozik, L. R. Schaeffer, and J.C.M. Dekkers. 1995a. Estimation of genetic parameters for test day somatic cell score. *J. Dairy Sci.* (accepted).
- Reents, R., J.C.M. Dekkers, and L. R. Schaeffer. 1995b. Genetic evaluation for somatic cell score with a multiple lactation test day model. *J. Dairy Sci.* (accepted).
- Schaeffer, L. R., and B. W. Kennedy. 1986. Computing strategies for solving mixed model equations. *J. Dairy Sci.* 69:575.
- Schneeberger, M., S.A. Barwick, G. H. Crow, and K. Hammond. 1991. Economic indices using breeding values predicted by BLUP. *J. Anim. Breed. Genet.* 109:180
- Swalve, H.H.. 1995. The effect of test day models on the estimation of genetic parameters and breeding values for dairy yield traits. *J. Dairy Sci.* 78:929.
- Tier, B., and H. U. Graser. 1991. Predicting breeding values using an implicit representation of the mixed model equations for a multiple trait animal model. *J. Anim. Breed. Genet.* 108:81.

TABLES

TABLE 1. Number of test-day records, cows, and levels of herd-test-date effects.

		Lactation			Total
		1	2	3	
Dataset 1 CANADA	Number of test day records	3,081,676	1,633,087	790,909	5,505,672
	Number of herd-test-date levels	377,573	292,213	203,924	
	Cows with records*	392,487	213,259	107,661	
Dataset 2 GERMANY	Number of test day records	10,117,929	7,298,206	5,201,018	22,617,153
	Number of herd-test-date levels	1,576,243	1,595,682	1,595,682	
	Cows with records*	1,487,279	865,284	629,283	

* Number of cows for lactation 2 and 3 are subsets of cows in lactation 1

TABLE 2. Variance components used for genetic eval. of SCS with models TD1 and TD2.

Random effect	Variances by lactation			Covariances between pairs of lactations		
	1	2	3	1 : 2	1 : 3	2 : 3
CANADA						
<u>SCS</u>	.202	.222	.319	.187	.200	.251
Animal additive genetic						
Permanent environment	.921	1.001	1.225	.276	.202	.504
Residual	1.054	1.213	1.269	0	0	0
<u>Protein Yield, (kg * 100)</u>						
Animal additive genetic	.398	.676	.667	.474	.443	.634
Permanent environment	.569	1.098	1.272	.417	.324	.667
Residual	.635	1.173	1.365	0	0	0
GERMANY						
<u>SCS</u>						
Animal additive genetic	.154	.270	.323	.183	.189	.286
Permanent environment	.737	.695	.764	.248	.209	.327
Residual	1.053	1.098	1.159	0	0	0

TABLE 3. Convergence criteria at various rounds of iteration for test-day model TD1 for genetic evaluation of SCS (dataset 1, correlations and deviations only for cows with observations).

Iteration	Log ₁₀ of relative difference C _s		Correlation of EBV with EBV at round 500		Absolute deviation of EBV with EBV at round 500			
					Mean		Maximum	
	Perm. envir.	EBV	Lact. 1	Lact. 3	Lact. 1	Lact. 3	Lact. 1	Lact. 3
20	-2.398	-1.422
50	-3.138	-2.897	.995	.994	.020	.051	.148	.315
100	-3.742	-3.442	.999	.999	.009	.025	.062	.166
150	-4.061	-3.746
200	-4.328	-4.000	1.000	1.000	.003	.009	.022	.080
250	-4.536	-4.205
300	-4.681	-4.365	1.000	1.000	.001	.005	.009	.045
400	-4.831	-4.559	1.000	1.000	.000	.002	.003	.019
500	-4.896	-4.697	-	-

TABLE 4. Correlations between estimated breeding values (dataset 1) from test-day model TD2 (combined EBV for lactations 1, 2, and 3) and EBV from three lactation average models (REP1, REP2, and ST) for SCS and protein yield (PY).

EBV TD2		Number of animals	Estimated breeding values from model					
			REP1		REP2		ST	
			SCS	PY	SCS	PY	SCS	PY
Bulls	No. of daughters							
	5 - 25	2598	.931	.728	.890	.712	.745	.688
	26 - 50	1137	.954	.780	.917	.771	.814	.744
	51 - 100	459	.968	.814	.940	.802	.807	.757
	> 100	320	.972	.839	.961	.813	.858	.766
Cows with observations		380,178	.957	.784	.940	.761	.815	.740

TABLE 5. Ranking of bulls with highest estimated breeding values (dataset 1) for SCS (most undesirable) and protein yield (PY) based on model TD2 (combined evaluation over lactations 1, 2, and 3) compared to the ranking based on estimated breeding values from lactation average models.

Number of daughters included in EBV	Number of bulls	Ranking for TD2	No. of bulls in common with TD2					
			REP1		REP2		ST	
			SCS	PY	SCS	PY	SCS	PY
5 - 25	2598	Top 20	14	11	11	9	5	6
		Top 50	34	30	30	17	20	18
		Top 100	70	60	60	47	44	40
26 - 50	1137	Top 20	15	11	11	7	11	5
		Top 50	38	37	37	24	22	23
		Top 100	80	76	76	51	55	51
51 - 100	459	Top 20	17	17	17	11	12	8
		Top 50	42	39	39	32	27	29
		Top 100	87	82	82	62	70	61
> 100	320	Top 20	15	15	15	14	11	8
		Top 50	42	41	41	33	34	32
		Top 100	89	89	89	75	73	69