Impact of Application of a Test Day Model to Dairy Production Traits on Genetic Evaluations of Cows

R. Reents, L. Dopp, M. Schmutz and F. Reinhardt
Vereinigte Informationssysteme Tierhaltung w.V. (VIT), Heideweg 1, D-27283 Verden,
Germany, Internet: www.vit.de

Abstract

A multiple lactation test day model was applied to predict genetic merit for dairy production traits in German 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. An adjustment procedure to account for within herd heterogeneous variances is described, considering a) number of contemporary records and b) production level on the particular test day, c) parity, and d) stage of lactation. Method of standardization for heterogeneous HTD variances has little impact on overall cow and bull rankings, but significant effects on EBVs for individual animals. Method of standardization has an impact on SD of EBVs and also on genetic trend estimates.

Introduction

A test day (TD) model which considers test day records as repeated observations within parity and as different traits across lactations is used since December 1995 in Germany in order to estimate intermediate, inofficial evaluations for Holstein dairy cattle twice a year. This multiple lactation test-day model (MLTDM) includes both random animal additive genetic and permanent environmental effects as different traits for each parity (Reents et al., 1995*a*,*b*). Fixed regressions (Ali and Schaeffer, 1987), which are nested within age and season of calving, region, parity, and calving interval are used to account for fixed environmental effects and for the shape of the lactation curve in order to define the management group as herd-test-day (HTD) effect instead of a herd-year-season classification. A herd-test-day classification allows to account for short term environmental variation from one month of testing to the next which is a very important advantage of a test-day-model compared to a lactation model (e.g. Reents, 1996; Wiggans and Goddard, 1997). Heterogeneous variances on a specific herd-test-day can be, among other reasons, due to a) different management of cows and b) different stage of lactation of the cows contributing test day records to the specific herd-test-day class. Therefore adjustment for heterogeneous within herd variance is complicated in a test-day model compared to a lactation model. A preadjustment method for heterogeneous HTD-variance is described. Further, comparisons of test day evaluations to the current official evaluations with a lactation model are made.

Materials and Methods

Data

Data consisted of test day records since January 1990 from the database maintained at Vereinigte Informationssysteme Tierhaltung w.V. (VIT), Verden, Germany (which contains about 75% of all Holstein cows in Germany) and other milk recording agencies. Following edits (see Reents et al., 1997) for the December 1997 evaluation 77,238,007 records from lactation 1 to 3 on 4,964,155 cows remained. Pedigree was completed for cows with identification of dam and maternal grandsire from the national pedigree file maintained at VIT. 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). The total number of animals evaluated was 7,302,167.

Model

For genetic evaluation of test-day observations, a multiple trait test day model (MLTDM) with repeated observations within each lactation was used.

The statistical model for analysis of test-day records was:

$$\begin{aligned} y_{ijkmn} &= HTD_{im} + A_{jm} + P_{jm} + RASC_{km} \\ &\quad + b_{km1}(D/c) + b_{km2}(D/c)^2 \\ &\quad + b_{km3}1n(c/D) + b_{km4}\left[1n(c/D)\right]^2 + e_{ijkmn} \end{aligned}$$

where

y_{ijkmn}	is the nth test day observation of the jth 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 jth cow in parity m (random)
$RASC_{km}$	is a region x age_of_calving x season_of_calving x calving_interval
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 945 lactation x RASC groups, resulting from: 3 lactations, 3 regions, 5 age_of_calving groups, 3 season_of_calving groups (Jan-Mar; Apr-Aug; Sep-Dec), 7 groups for calving_interval (< 320 days, 321-350 days, ..., > 470 days). According to Reents et al. 1995*b* 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. The statistical model for the official genetic evaluation with the multi lactation model (MLM) is:

$$y_{ijm} = HYS_{im} + a_{jm} + e_{ijm}$$

where

y_{ijm}	is the yield of cow j in part-lactation m (three 100 day parts from first lactation, the second lactation and the third lactation are considered as 5 genetically different traits)
HYS_{im}	is a fixed herd-year-season effect
a_{jm}	is a random additive genetic effect of animal j, and
e_{ijm}	is a random residual effect.

Variance components used for MLTDM were as in Reents et al. (1995*a*) and for MLM as used in the official evaluation in Germany. In the tables of this paper MLM-1 denotes results from the official evaluation in August 1997 (starting with calvings from 1979), MLM-2 is a run which uses similar data as in the MLTDM (only with calvings since 1990).

Standardization for within herd heterogeneous variances

Selection among groups with heterogeneous variances is a problem in selection schemes (Hill, 1984). For the MLTDM a method to standardize heterogeneous variances within herd test days (hetHTD_V) was developed similar to the method described by Wiggans and VanRaden (1991) and used in the official evaluation (MLM) in Germany. The applied standardization method accounts for heterogenity of the phenotypic variance and is performed as a preadjustment for each yield trait separately. In the first step all TD records are corrected for stage of lactation and RASC effects (region, age, season, calving interval) using lactation curves estimated from an LSQ-Model (fixed part of the MLTDM). Since scale effects should be preserved, the average standard deviation of all herd test days within the same production level is assumed as the "expected" value. The production levels are classified according to the corrected HTD-mean, i.e. for milk yield with an interval of 1,0 kg in the range of 10-30 kg and 10-35 kg for the first and later lactations, respectively. For fat yield range is from 0.25 kg to 1.5 kg (1.75 kg for later lactations) with a spacing of 0.05 kg. For protein yield range is from 0.2 kg to 1.0 kg (1.2 kg for later lactations) with a spacing of 0.033 kg. The observed standard deviation for each herd test day (s_{obs}) is combined with the "expected" standard deviation according to the production level (s_{pl}) to get the weighted standard deviation (s_{wgt}) depending on herd size (n): $s_{wgt} = (x * s_{pl} + n * s_{obs})/(x+n).$ Two empirical values for x were examined, a value of x = 10 puts more weight on the observed standard deviation of the specific herd-test-day, whereas a value of x = 20 shifts the weighted standard deviation more to the average SD of the specific production level. Then the weighted standard deviation is used to calculate the standardization factor for each herd test day (sf): $sf = s_{pl} / s_{wgt}$, thus with x = 10 standardization factors sf have a larger magnitude compared to x = 20. In a second step, the standardization is applied to the corrected (stage of lactation within RASC) observations (y), resulting in $y = \overline{y} + sf * (y - \overline{y}).$ After the standardization y_s has to be adjusted for the RASC effect again to avoid double counting, because RASC effects are estimated in the final iteration process again. This standardization method provides constant standardization factors of ancient test days on consecutive estimation runs. To compare the effect of standardization on MLTDM results three different runs with data from June 1997 were performed:

- MLTDM - 1 = heterogeneous variances not considered
- MLTDM - 2 = heterogeneous variances considered (x = 10)
- MLTDM - 3 = heterogeneous variances considered (x = 20)

Computational aspects

Application of test day models under an animal model leads to a dramatic increase in computational demand. First nearly ten times more records (with a four week interval between samples) have to be processed compared to one lactation measure and second models are more complicated because a proper adjustment for stage of lactation has to be defined in the statistical model. Also consideration of herd-test-day classes leads to much more levels compared to herd-year-season classification. German Holstein population consists of about 2.5 Mio cows under milk recording, from which about 1.9 Mio cows contribute nearly 20 Mio new test day records from lactations 1 to 3 per year. Due to the implicit representation of the mixed model equations (Reents et al., 1995*b*) the multi lactation test day model for the described data set, comprising of 7.3 Mio animals with 77 Mio test day records, required less than 650 MBytes of memory if iteration on data was performed reading a file from disk in each round of iteration. Then CPU time per round of iteration is about 15 minutes on a HP 9000/K420 Unix workstation with 2 GBytes of random access memory (RAM). Storing all test day records in RAM reduced time per round of iteration to 5.5 minutes CPU time with an increase in demand of RAM to 2 GB. For comparison a run with the official lactation model (starting with calvings from 1990) would require 800 MB of RAM and 2 minutes CPU time per round of iteration. In the described MLTDM every cow with TD records contributes 6 equations (3 for EBVs and 3 for permanent environmental effects in lactations 1, 2, and 3) and every ancestor 3 equations for EBVs in lactations 1, 2, and 3 to the mixed model equations.

Application of a random regression model as proposed by Jamrozik et al., 1997 (assuming Wilmink=s function for the shape of the lactation curve), would lead to 36 equations for each animal without performance records and to 72 equations for animals with performance records. Therefore application of this model to the large German Holstein population would lead to a dramatic increase in computing power required.

Results and Discussion

EBVs from MLTDM are on a per day production scale. EBVs from MLTDM can be transformed to a 305 day equivalent by multipling with 305 as proposed by Ptak and Schaeffer (1993). Table 1 displays standard deviations of estimated breeding values for cows from different runs. Without any standardization for heterogeneous herd-test-day variances (MLTDM-1) standard deviation (SD) of cow EBVs is much higher (554 kg milk) than SD from the official lactation model MLM-1 (507). Applying a strong adjustment method (MLTDM-2, x = 10) reduces SD of cow EBVs dramatically to a SD of 471 kg milk. Applying a less strict adjustment procedure (MLTDM-3, x=20) results in a SD of 514 kg milk, which is very similar to the figure from the current official evaluation with MLM-1. Table 1 also shows that method of adjustment for hetHTD_V has little impact on overall ranking of cows in all three test-day-model runs (r= .99). However, for rankings of individual (TOP) cows an adjustement for hetHTD_V is necessary to avoid overrepresentation of cows with preferential treatment in national TOP-cow lists and also to avoid bias in EBVs from bulls with a second country proof in Germany. Correlations between EBVs from TD models and the lactation models MLM-1 and MLM-2 is as high as .86 to .87.

Table 2 displays SD and correlations for EBVs of AI bulls from different models. Same as for cows method of standardization has little impact on ranking of all bulls (r=.99) but for individual bulls method of standardization can have a significant impact on bull EBV. Comparison of MLM-1 and MLM-2 indicates that use of milk recording data since 1990 instead of using data since 1979 seems to be of little impact on ranking of the last 6 years of AI bulls, as the very high agreement (r=.99) shows. Figures 1 and 2 display genetic trend estimates from the five models investigated for bulls and cows. Genetic trend is significantly higher from all three test-day models in comparison to the lactation models. Highest genetic trend is found with MLTDM-1, which is without any adjustment for hetHTD_V. Smallest genetic trend among the test day models studied is found with MLTDM-2, which is with a strict adjustment for heterogeneous variances. Genetic trend for MLTDM-3 (less strict adjustment) is intermediate to MLTDM-1 and MLTDM-2.

Tables 3 and 4 show, that due to the higher genetic trend in test-day-models more young animals are represented in TOP-100 bull rankings and also in TOP-1000 cow lists.

Conclusions

Application of a test day model for dairy production traits is feasible even on a large scale national dataset. Ranking from bull EBVs compared to a lactation model change substantially, i.e. correlations between MLTDM results and MLM are in a range of .94 and for cows in a range of .87. Method of standardization for heterogeneous HTD variances has little impact on overall cow and bull rankings, but significant effects on EBVs for individual animals could be found. Method of standardization has an impact on SD of EBVs and also on genetic trend estimates. Likely method MLTDM-3 will be chosen for official implementation of a TD model in Germany.

References

Ali, T.E. & Schaeffer, L.R. 1987. Accounting for covariances among test day milk yields in dairy cows. *Can. J. Anim. Sci. 67*, 637.

Hill, W.G. 1984. On selection among groups with heterogeneous variance. *Anim. Prod. 36*, 59.

Jamrozik, J., Schaeffer, L.R., Liu, Z. & Jansen, G. 1997. Multiple trait random regression test day model for production traits. *INTERBULL Bulletin No. 16*, 1997 .

Ptak, E. & Schaeffer, L.R. 1993. Use of test day yields for genetic evaluation of dairy sires and cows. *Livest. Prod. Sci. 34*, 23.

Reents, R. 1996. The use of test day models in genetic evaluation. *INTERBULL Bulletin No. 12*, 1996.

Reents, R., Jamrozik, J., Schaeffer, L.R. & Dekkers, J.C.M. 1995*a*. Estimation of genetic parameters for test day records of somatic cell score. *J. Dairy Sci. 78*, 2847.

Reents, R., Dekkers, J.C.M. & Schaeffer, L.R. 1995*b*. Genetic evaluation for somatic cell score with a test day model for multiple lactations. *J. Dairy Sci. 78*, 2858.

Reents, R., Dopp, L. & Reinhardt, F. 1997. Experiences with a genetic evaluation using test day data. *INTERBULL Bulletin No. 16*, 1997.

Wiggans, G.R. & Goddard, M.E. 1997. A computational feasible test day model for genetic evaluation of yield traits in the United States. *J. Dairy Sci. 80*, 1795.

Wiggans, G.R. & VanRaden, P.M. 1991. Method and effect of adjustmend for heterogeneous Variance. *J. Dairy Sci. 74*, 4350.

Table 1. Standard deviations (SD) and correlations (x100) between cow EBVs from different models / estimations, milk yield, cows born from 1987 to 1995.

Model	N	SD	MLTDM -2	MLTDM -3	MLM -1	MLM -2
MLTDM -1	3,770,791	554	99	99	86	87
MLTDM -2	3,770,791	471	-	99	86	86
MLTDM -3	3,770,791	514		-	87	87
MLM -1	4,343,732	515			-	94
MLM -2	3,044,472	507				-

Table 2. Standard deviations (SD) and correlations (x100) between bull EBVs from different models / estimations, milk yield, bulls born from 1987 to 1992.

Model	N	SD	MLTDM -2	MLTDM -3	MLM -1	MLM -2
MLTDM -1	3168	684	99	99	94	94
MLTDM -2	3168	574	-	99	94	94
MLTDM -3	3168	622		-	94	94
MLM -1	3118	540			-	99
MLM -2	3146	545				-

Table 3. Distribution of birthyears of bulls within TOP-100-rankings within each model (bulls born from 1987 to 1992).

Model	1992	1991	1990	1989	1988	1987
MLTDM -1	49	32	7	3	4	5
MLTDM -2	48	30	9	4	5	4
MLTDM -3	49	29	9	4	4	5
MLM -1	41	26	10	13	6	4
MLM -2	43	26	10	11	6	4

Table 4. Distribution of birthyears of cows within TOP-1000-rankings within each model, cows born from 1988 to 1995.

Model	1995	1994	1993	1992	1991	1990	1989	1988	av. Birthy.
MLTDM -1	16	411	339	155	55	13	9	2	931
MLTDM -2	19	400	355	158	53	8	7		93.1
MLTDM -3	19	390	360	162	52	9	8		93.1
MLM -1	19	209	206	226	154	79	73	34	92.0
MLM -2	22	288	238	221	130	51	33	17	92.5

Figures

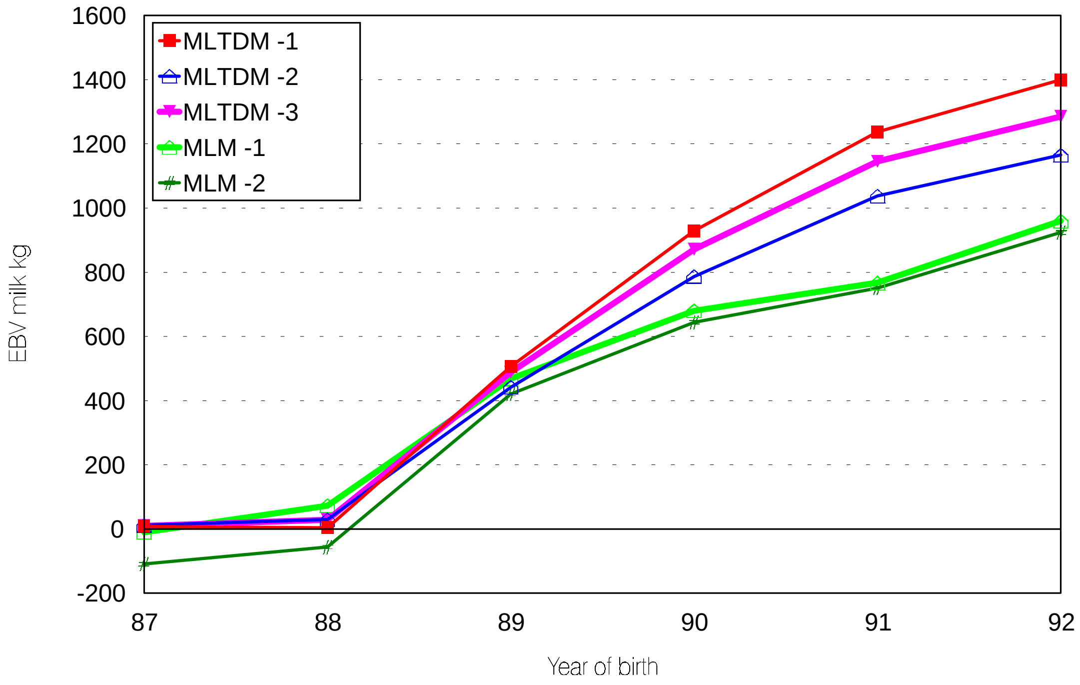

Figure 1: Genetic trend for bulls, born from 1987 to 1992, N = 3168

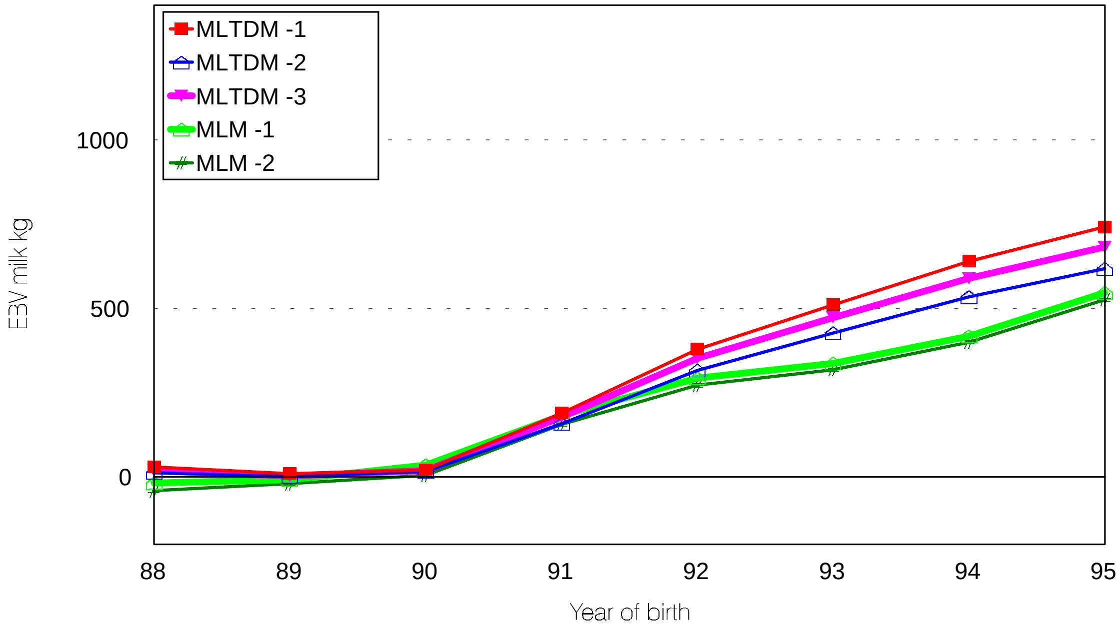

Figure 2: Genetic trend for milk yield of cows, born from 1988 to 1995, N= 3.044.472

1