Adjustment for Heterogeneous Herd-Test-Day Variances

G. J. Kistemaker and L.R. Schaeffer
Centre for Genetic Improvement of Livestock
University of Guelph, Guelph, Ontario, Canada

Currently lactation records are adjusted for within-herd heterogeneity of variance. Heterogeneity of variance can cause biases in estimated breeding values for cows producing in herds with variances that are above or below the average variance. Herds with high variances will have a high range of estimated breeding values and the best animals in these herds are overestimated. Correcting for heterogeneous herd variances reduces this problem.

Three different methods have been used to account for heterogeneous herd variances and these methods could be adapted for use in a test day model. Data can be adjusted before the evaluation to standardize the phenotypic variance. This method assumes that the heritability is the same for all herds and does not depend on the phenotypic variance. The second approach is to adjust for heterogeneous variance in the evaluation itself by increasing residual variances for herds with high phenotypic variances. This method assumes that higher phenotypic variances are due to higher residual variances and heritabilities for herds with higher variances are lower. There is no evidence of a decrease in heritability for high variance herds and this method might over-adjust the data. The third method is the joint estimation of breeding values and heterogeneous variances (Meuwissen et al., 1996). Joint estimation also assumes constant heritability but has the disadvantage of an increase in CPU time needed to perform the evaluation. Pre-adjusting the data was used in the current research because the CPU requirements are lower than using joint estimation, and pre-adjusting the data before the evaluation seems to be in better agreement with reported results than changing residual variances.

Pre-adjusting the data requires the estimation of herd-test-day variances for test day models. Estimated herd-test-day variances are used to adjust test day records to account for heterogeneous herd test day variances. The objectives of this research were to develop a method to estimate herd-test-day variances and to adjust test day records for heterogeneous herd test day variances.

Adjustment for heterogeneous herd-test-day variances consists of three steps:

1. Estimation of variances on herd-test-days,
2. Calculation of adjustment factors for herd test days, and
3. Use of adjustment factors to account for heterogeneity of herd test day variances.

Estimation of variances on individual herd-test-days

In order to obtain accurate estimates of test day variances, observations have to be adjusted for fixed effects such as days in milk, lactation number, and season of calving. For example, if milk yields are not adjusted for effects of days in milk then herds with seasonal calving will have lower test day variances because all cows in these herds are in approximately the same stage of production on each test day. Data was adjusted for fixed effects using the following model:

- Effect of days in milk
- Effect of herd test day
- Wilmink curve for each age-season-parity class
- Wilmink curve for each herd

Variances of test day records depend on the number of days in milk, these differences in variance have to be removed in order to obtain accurate estimates of herd test day variances. Residuals from the previous model were used to estimate the residual standard deviation (σ) and standard deviations for each day in milk (σ_{DIM}). After obtaining these standard deviations each residual (R) was adjusted for DIM variances using the following model:

$$(R_{\text{standardized}}) = \frac{R}{(\sigma_{DIM})} \quad \sigma$$

Standardized residuals $(R_{\text{standardized}})$ were used to calculate herd-test-day variances.

Calculation of adjustment factors for herd-test-days

Adjustment factors for herd-test-days can be calculated directly by dividing estimated standard deviations by the average standard deviation. However, information from previous test days can be used to improve estimates of test day standard deviations if standard deviations of test days in the same herd are correlated. The number of days in milk between test days varies and this variation should be taken into account when using information from previous test days.

Correlations between test day standard deviations (within the same herd) were calculated separately for different numbers of days between tests. Only test day standard deviations based on at least 20 observations were used and correlations were calculated if at least 10 pairs of test day standard deviations were available. A line was fitted through these correlations using the following model:

$$C_t = a + bc + dt$$

where

C_t is the correlation between test days which are t days apart, and
a, b, c and d are parameters

Estimated standard deviations in small herds are based on relatively few observations and show high variations from one test day to the next. Extreme adjustment factors can be avoided by regressing those factors towards the standard deviation of all test day records. This would be more important in small herds were little information about the standard deviation is available. Adjustment factors for each herd test day were calculated as:

$$Factor = \frac{W\,\sigma_{mean} + \sum_{i=0}^{5}(C_i df_i \sigma_i)}{\left[W + \sum_{i=0}^{5}(C_i df_i)\,\sigma_{mean}\right]}$$

where

W is the weight used for the mean standard deviation

σ_{mean} is the standard deviation of all test day records, i indicates the test day, i=0 is the current test day, i=1 is the previous test day, Y , and i=5 is five test days earlier

C_i is the estimated correlation between the standard deviation on the current test day and the standard deviation on test day i, if t > 180 then C_i was set to 0

df_i is the degrees of freedom on test day i and is equal to the number of observation minus one, and

σ_i is the estimated standard deviation on test day i

Using adjustment factors to account for heterogeneous variances

Test day observations were adjusted using the following equation:

$$Y_{adjusted} = (Y - Y_{predicted})/\text{factor} + Y_{predicted}$$

where

$Y_{adjusted}$ is the adjusted observation used for the genetic evaluation

Y is the test day observation before adjustment, factor is the adjustment factor and

$Y_{predicted}$ is the predicted yield obtained from the model with fixed effects

Application to Jersey test day observations

Test day milk, fat and protein yields and somatic cell score (SCS) of Jersey cows were used for testing the effect of different methods to adjust for heterogeneous variances. All Jersey test day observations from herds with at least five test days, and test days with at least 5 observations were selected. This data set contained observations from ten years produced in 442 different herds. In total 33,524 animals were included in the analysis. Weights used to calculate adjustment factors were selected to give equal weight to the overall standard deviation and herd test day standard deviations in herds with 0 (W=0), 5 (W=15), 10 (W=35) and 15 (W=55) cows which are tested every 30 days. These weights (W) were obtained using correlations for protein yield and were used for all four traits. Adjustment factors were calculated for each herd-test-day and each trait using the four weights. Four data sets with adjusted observations were obtained using adjustment factors based on different weights, in addition the unadjusted observations were used. A multiple lactation random regression test day model was fitted to each data set and trait separately. Breeding values for 305-day milk yields were estimated for the first three lactations. Estimated 305-day breeding values for the first three lactations were averaged to give one estimated breeding value per animal. This average 305-day breeding value was used for the comparison of the different adjustment factors.

Results

Correlations between standard deviations on test days (within the same herd) are shown in Figure 1. Correlations between test days close together were high and this correlation decreased as the interval between tests became longer. Correlations for the yield traits (milk, fat and protein) were very similar and these correlations were higher than correlations for SCS.

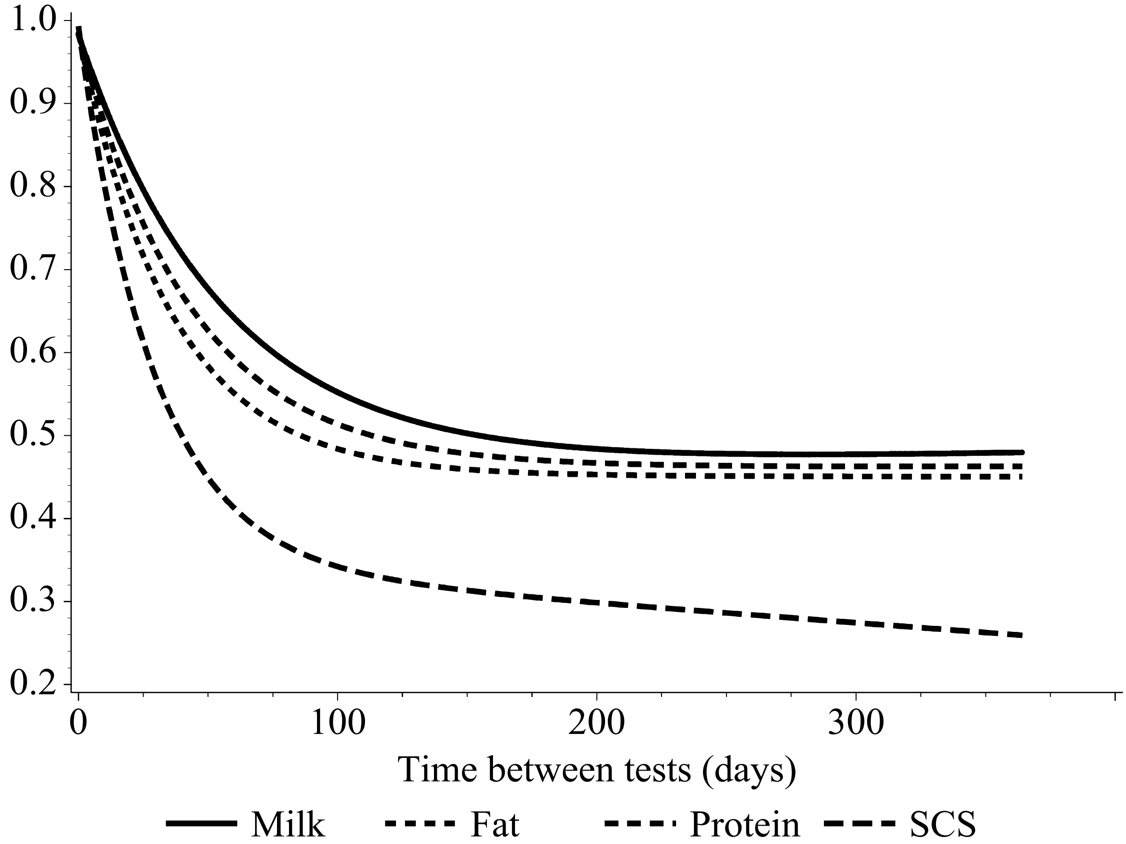

Figure 1. Correlations between test day standard deviations within a herd, for different traits and different lengths of the interval between test days.

Correlations between breeding values estimated from unadjusted data and breeding values estimated from adjusted data ranged between .985 and .995 (Table 1). Correlations were very similar for the yield traits (milk, fat and protein) but were higher for SCS. Correlations increased when an adjustment factor with a higher weight was used, these adjustment factors were regressed more towards the mean and had a lower influence.

Table 1. Correlations between EBV without HV adjustment and EBV with HV adjustment

Weight on mean	Milk, Fat	Prot	SCS
No weight	0.985	0.984	0.993
5 cows	0.991	0.991	0.997
10 cows	0.994	0.994	0.998
15 cows	0.995	0.995	0.999

The main problem of heterogeneous herd variances is the selection of bull dams. To see the impact of the different adjustments on selection of bull dams, the best 250 cows (approximately 1 %) with test day observations were selected from each model. Using adjustment factors with a weight of zero resulted in the lowest number of cows in common (Table 2) and thus the highest number of rerankings. The percentage reranking was highest for milk yield (23 to 39%) and lowest for SCS (11 to 23%).

Table 2. Number of top 250 cows (1%) from unadjusted data still in top 250 after adjustment

Weight on mean	Milk	Fat	Prot	SCS
No weight	153	169	167	194
5 cows	171	186	188	207
10 cows	186	197	197	215
15 cows	191	205	203	221

Adjustment for heterogeneity of variance reduces the difference between estimated breeding value (EBV) and the parent average (PA) for progeny tested sires, the variance of (EBV - PA) is also reduced (Meuwissen et al., 1996, J. Dairy Sci. 79:310). Therefore, (EBV - PA) was calculated for each progeny tested sire (> 49 progeny), the mean and variance of (EBV - PA) was calculated from all 80 progeny tested sires in the data set. From Tables 3 and 4 it can be clearly seen that the adjustments decrease both the mean and the variance of (EBV - PA) and the adjustments seem to be effective. Adjustment factors with a higher weight resulted in a lower reduction of the mean and variance of (EBV - PA). Again the effect of the adjustment was the lowest for SCS.

Table 3. Mean and variance of EBV - PA for Milk and SCS (80 Jersey bulls)

	Milk		SCS	
Method	Mean	Var	Mean	Var
Unadjusted	**-158**	**123**	**23.5**	**6,850e+19**
No weight	-143	108	22.4	
5 cows	-145	112	23.0	
10 cows	-147	114	23.1	
15 cows	-148	115	23.2	

Table 4. Mean and variance of EBV - PA for Fat and Protein (80 Jersey bulls)

	Fat		Protein	
Method	Mean	Var	Mean	Var
Unadjusted	**-6.16**	**267**	**-5.34**	**158**
No weight	-5.33	222	-4.68	139
5 cows	-5.50	236	-4.81	144
10 cows	-5.62	242	-4.90	146
15 cows	-5.69	246	-4.96	148

Conclusion

Results show that the adjustment for heterogeneous variances reduced the mean and variance of (EBV - PA) indicating that the adjustment was effective. Correlations between estimated breeding values were high but adjustment had a major effect on the selection of bull dams. Up to 39% of the bull dams selected from the unadjusted data were replaced when adjusting data for heterogeneous variances. Using a low weight on the overall standard deviation resulted in the most effective adjustment.

References

Meuwissen T.H.E., De Jong, G. & Engel, B. 1996. Joint estimation of breeding values and heterogeneous variances of large data files. *J. Dairy Sci. 79*, 310-316.