

# On the Use of Projected 305-day Yields in the Breeding Value Prediction for Production Traits in the Netherlands

*Sijne van der Beek and Gerben de Jong*  
NRS, P.O. Box 454, 6800 AL Arnhem, The Netherlands  
E-mail: beek@nrs.nl, jong@nrs.nl

---

## Abstract

Projected 305 day yield has less phenotypic variance than realized 305 day yield, a genetic covariance with realized yield that is smaller than the genetic variance of realized yield, and a genetic correlation with realized yield that is below 1. The model used for breeding value estimation, however, assumes for projected yield that: it has a genetic covariance with realized yield that is equal to the genetic variance of realized yield, a residual variance that is larger than the residual variance of realized yield, and a genetic correlation with realized yield of 1. Most of the discrepancy between model and data can be reduced by: 1) rescaling projected yield so that it has a genetic covariance with realized yield that is equal to the genetic variance of realized yield, and 2) correcting projected yield of a cow for the breeding values for lactation curve parameters of the sire, which increases the genetic correlation between projected and realized yield.

---

## Introduction

In the Netherlands, a single-trait animal model is used to predict breeding values for milk production traits. Realized and projected 305-day yields are used. Until August 1997, projected yields were included if based on a lactation of at least 60 days. However, inclusion of projected yields based on short lactations lead to unjustifiable dynamics of the estimated breeding values (ebv's). On average the ebv's of the better young bulls increased with increasing lactation length of the daughters. For proven bulls, on average ebv's decreased if the first batch of projected yields based on short lactations of second crop daughters came available. To circumvent this problem, currently only projected yields based on lactations longer than 180 days are included. In addition, research on adapting the breeding value estimation procedure was intensified. The results of this research are reported here. All results are for the trait protein yield.

## Theory

Van Raden et al. (1991) showed that the relation between lactation length and ebv's is due to a discrepancy between the true covariance structure of the data included in the genetic evaluation, and the covariance structure implied by the method and model used. The model assumes that the genetic correlation between projected and realized yield is 1, the genetic covariance between projected and realized yield is equal to the genetic variance of realized yield, the heritability of projected yield is smaller than the heritability of realized yield, and consequently that the phenotypic and error variance for projected yield are larger than that of realized yield. In reality, however, projected yield has less phenotypic variance than realized yield and a genetic correlation smaller than 1 (Wilmink, 1987a). Expansion, or rescaling, of projected yields reduces the discrepancy between model and data (Van Raden et. al., 1991).

Expansion removes differences between implied and true genetic covariance between projected and realized yield, as well as the differences between true and implied residual variance, but does not change the genetic correlation between projected and realized yield. Thus a discrepancy between implied and true genetic correlation remains.

### Data

Protein records of heifers starting lactation after 1-1-1993 were used. Data were used from 168355 heifers originating from 1531 black and white sires having at least 10 daughters in the data. Projected 305 day yields were computed from the test day data in those records. Per lactation record 9 projected 305 day yields were computed:  $p_{60}$ ,  $p_{90}$ ,  $p_{120}$ ,  $p_{150}$ ,  $p_{180}$ ,  $p_{210}$ ,  $p_{240}$ ,  $p_{270}$  and  $p_{300}$ , where  $p_x$  is projected 305 day yield using all test days until day  $x$ . Projected yield  $p_{300}$  was assumed to be equivalent to realized 305 day yield. Projected and realized 305 day yields were corrected for age at calving using multiplicative correction factors.

### Estimation of covariance components

REML estimates were computed for the covariances between the nine projected 305 day yields. Estimates were for a sire model:

$$p_{xijk} = HYS_i + s_j + e_{ijk}$$

where

- $p_{xijk}$  is the projected 305 day yield using all test days until day  $x$  for cow  $jk$  in herd-year-season  $i$
- $HYS_i$  is fixed effect of herd-year-season
- $s_j$  is the random additive genetic effect of the sire, and
- $e_{ijk}$  is a random residual effect.

### Estimation of breeding values for lactation curve parameters

De Jong (1997) estimated breeding values for lactation curve parameters. A Wilmink function (Wilmink, 1987b) was fitted for each cow using raw test day data:

$$y_x = \beta_0 + \beta_1 x + \beta_2 e^{-0.05x}$$

where

- $y_x$  is the production on test day  $x$
- $\beta_{0,1,2}$  are the lactation curve parameters that are estimated for each cow.

The estimated  $\beta_1$ 's and  $\beta_2$ 's were then used as data for the estimation of breeding values for this lactation curve parameter:

$$\beta_{1ijklm} = HYS_i + age_j + m_k + s_j + e_{ijklm}$$

where

- $age_j$  is fixed effect of age at calving
- $m_k$  is fixed effect of calving month
- $s_j$  is the random additive genetic effect of the sire, and
- $e_{ijklm}$  is the random residual effect.

The heritability of the sire effect was set to 0.15 (Galesloot, 1996).

### Rescaling of projected yields

Projected yields were rescaled using:

$$r_{xijk} = r_{f_x} \times (p_{xijk} - \bar{p}_{p_{xi}}) + \bar{p}_{p_{xi}}$$

where

- $r_{xijk}$  is the rescaled projected yield
- $r_{f_x}$  is the rescaling factor for projected yield
- $\bar{p}_{p_{xi}}$  is the average of the projected yields in herd-year-season class  $i$ .

Rescaling factor  $r_{f_x}$  is computed as:

$$r_{f_x} = \sigma_g^2(p_{300}) / \sigma_g(p_x, p_{300})$$

where

- $\sigma_g^2(p_{300})$  is the genetic variance for  $p_{300}$ , and
- $\sigma_g(p_x, p_{300})$  is the genetic covariance between  $p_x$  and  $p_{300}$ .

### ***Analysis of the difference between projected and realized 305 day yield***

Differences  $d_x$  between  $p_x$  and  $p_{300}$  and  $\delta_x$  between  $r_x$  and  $p_{300}$  were computed and averaged over sires. Differences were regressed on breeding values for the lactation curve parameters  $\beta_{1,2}$  and the breeding value for 305 day yield.

## **Results**

### ***Projected yields***

The genetic covariance and correlation between  $p_x$  and  $p_{300}$ , and the heritability and phenotypic variance of  $p_x$  increase with increasing lactation length (Table 1). Projected yield has less phenotypic and genetic variance than realized yield ( $p_{300}$  is realized yield). At 210 days the genetic correlation between realized and projected yield approaches 1, but the variance of projected yield at 210 days is only 90% of the variance in realized yield.

### ***Difference between projected and realized yield***

The difference between projected and realized yield is related to the breeding value for 305 day yield, especially at the beginning of the lactation (Table 2). The breeding value for  $\beta_1$  explains most variance between day 100 and day 200, whereas the breeding value for  $\beta_2$  explains some variance before day 60, i.e. before the peak of the lactation curve.

### ***Rescaled projected yield***

A comparison between Table 1 and Table 3 shows that genetic covariance and phenotypic variance increase due to rescaling, whereas heritability and genetic correlation are unaffected. Genetic covariance between  $r_{60}$  and  $p_{300}$  is higher than genetic variance of  $p_{300}$ , for  $r_{90}$  to  $r_{270}$  the genetic covariance is very close to the genetic variance of  $p_{300}$ . The phenotypic and residual variance of rescaled projected yield are largest for  $r_{60}$  and reduce with increasing lactation length.

### ***Difference between rescaled and realized yield***

Rescaling reduces the effect of the breeding value for 305 day yield on the difference between projected and realized yield (Table 4). The breeding value for  $\beta_1$  explains much more variance than the other breeding values. The breeding value for  $\beta_2$  explains variance at the begin of the lactation. At the begin of the lactation there seems to be interaction between the breeding values for  $\beta_1$  and  $\beta_2$  since they jointly explain more variance than the sum of the variances explained by either variable.

Results in Table 4 are from an analysis including all sires irrespective of the reliability of their breeding values. In Table 5 are the results of an analysis including 151 sires that have breeding values with a reliability of above 95%.

Table 5 shows that more than 50% of the differences between projected and realized yields can be explained by a reliable breeding value for  $\beta_1$ .

### ***Projected yields that are rescaled and corrected for the breeding value for $\beta_1$***

Results from the regression of  $\delta_x$  on the breeding value for  $\beta_1$  were used to correct rescaled yields. Table 6 shows that this correction increases the genetic correlation between projected yield and realized yield. The genetic correlation is 1.00 at day 120 and 0.99 at day 90. Correction increased the heritability, which is a logical consequence of adding a value to the rescaled projected yield of all daughters of a sire. Compared to rescaled yields  $r_x$ , the genetic covariance between  $z_x$  and  $p_{300}$  is increased. The genetic covariance between  $z_x$  and  $p_{300}$  is now clearly above the genetic variance of  $p_{300}$ .

## **Discussion**

Rescaling projected yields as suggested by Van Raden et al. (1991) reduces the relation between lactation length and the ebv for 305 day yield. Our rescaling factor was the genetic variance of realized yield divided by the genetic covariance between realized and projected yield whereas Van Raden et al. used the ratio of genetic standard deviation of realized yield over genetic standard deviation of

projected yield. Rescaling factors based on genetic covariances are larger than rescaling factors based on genetic standard deviations. When using rescaling factors based on genetic covariances, the model used for breeding value estimation properly describes the genetic covariances between rescaled yield and the breeding values of other animals, but assumes that rescaled yields have less genetic variance than they actually do. For instance, the genetic variance of  $r_{90}$  was 0.2369 (unrounded value)  $\times 1132 = 268$ , about 10% above genetic variance of  $p_{300}$ . So, ebv's of sires will be correct, whereas ebv's of cows with rescaled yields will show too much variance which means that on average the ebv's of the better animals will reduce as lactation proceeds. When rescaling factors based on genetic standard deviations are used, cow ebv's are correct but ebv's of sires have too little variance, i.e., the ebv's of the better half of the sires are underestimated when based on projected yields.

Cow ebv's are overestimated or sire ebv's underestimated when the genetic correlation between projected and realized yield is not 1. A genetic correlation under 1 causes a second problem: based on projected yields sires rank differently than based on realized yield. Correcting rescaled projected yield for the sire breeding value for  $\beta_1$  results increases genetic correlations. For projected yields based on a lactation of 90 days the genetic correlation was 0.99, and 120 days the correlation was one. Adding sire variance to projected yields increases the genetic covariance between projected and realized yield above the genetic variance of realized yield. For instance,  $z_{90}$  has a genetic covariance with  $p_{300}$  of 256 and a genetic variance of 278 (the unrounded  $h$  is 0.2449). The rescaling factor for  $p_{90}$  should be reduced by a factor 256/242.

This would reduce genetic covariance between  $z_{90}$  and  $p_{300}$  to 242 and also reduce variance of  $z_{90}$  to  $278 / (256/242) = 248$ , while the genetic variance of  $r_{90}$  was 268. Thus increasing the genetic correlation also solves the problem of too much genetic variance in cow ebv's.

Rescaling improves the ebv's of all animals. Correcting rescaled yield for breeding values of  $\beta_1$  (slope) will mainly help sires that already have daughters with completed records. Young bulls will still rerank due to differences in slope of the lactation curve. Also, correction of projected yield for breeding values for the slope does not reduce reranking due to differences in genetic merit for first and later lactations that are currently not accounted for in the model. For this, a multitrait animal model is required. In the Netherlands a multitrait model will be introduced as part of the test-day model under development.

## References

- Galesloot, P.J.B. 1996. De genetische relatie tussen persistentie van melkproductie en vruchtbaarheid bij gelijk productieniveau. *Msc thesis*, Wageningen Agricultural University.
- Jong, G.J. de. 1997. Persistentie index. NRS internal publication.
- VanRaden, P.M., Wiggans, G.R. & Ernst, C.A. 1991. *J. Dairy Sci.* 74, 4344-4349.
- Wilmink, J.B.M. 1987a. Efficiency of selection for different cumulative milk, fat and protein yields in first lactation. *Livest. Prod. Sci.* 17, 211-224.
- Wilmink, J.B.M. 1987b. Adjustment of test-day milk, fat and protein yield for age, season and stage of lactation. *Livest. Prod. Sci.* 16, 335-338.

Table 1. Genetic covariance ( $\sigma_g$ ) between  $p_x$  and  $p_{300}$ , genetic correlation ( $r_g$ ) between  $p_x$  and  $p_{300}$ , and heritability (h) and phenotypic variance ( $\sigma_p^2$ ) of  $p_x$  for protein yield.

	p <sub>60</sub>	p <sub>90</sub>	p <sub>120</sub>	p <sub>150</sub>	p <sub>180</sub>	p <sub>210</sub>	p <sub>240</sub>	p <sub>270</sub>	p <sub>300</sub>
$\sigma_g$	121	147	170	188	205	218	229	237	242
$r_g$	0.90	0.94	0.97	0.98	0.99	1.00	1.00	1.00	1.00
$\sigma_p^2$	353	427	484	541	595	639	678	702	716
h)	0.22	0.24	0.26	0.28	0.30	0.31	0.32	0.33	0.34

Table 2. Variance in daughter average for  $d_x$  explained by breeding value for 305 day production (ebv<sub>305</sub>), breeding value for  $\beta_1$ (ebv<sub>B1</sub>), breeding value for  $\beta_2$ (ebv<sub>B2</sub>) and combinations of those variables.

	ebv <sub>305</sub>	ebv <sub>B1</sub>	ebv <sub>B2</sub>	ebv <sub>305</sub> + ebv <sub>B1</sub>	ebv <sub>B1</sub> + ebv <sub>B2</sub>	ebv <sub>305</sub> + ebv <sub>B1</sub> + ebv <sub>B2</sub>
d <sub>60</sub>	34.1	14.3	7.9	41.9	29.1	48.4
d <sub>90</sub>	29.8	19.3	2.8	41.9	27.3	44.5
d <sub>120</sub>	25.1	20.7	0.5	38.9	24.3	39.5
d <sub>150</sub>	21.9	21.7	0.0	37.0	23.3	37.0
d <sub>180</sub>	19.1	21.3	0.0	34.3	22.0	34.2
d <sub>210</sub>	13.1	16.2	0.6	24.8	16.1	25.3
d <sub>240</sub>	8.4	7.7	0.4	13.7	7.7	14.1
d <sub>270</sub>	3.8	6.8	0.7	9.0	6.8	9.5

Table 3. Genetic covariance ( $\sigma_g$ ) between  $r_x$  and  $p_{300}$ , genetic correlation ( $r_g$ ) between  $r_x$  and  $p_{300}$ , and heritability (h) and phenotypic variance ( $\sigma_p^2$ ) of  $r_x$  for protein yield.

	r <sub>60</sub>	r <sub>90</sub>	r <sub>120</sub>	r <sub>150</sub>	r <sub>180</sub>	r <sub>210</sub>	r <sub>240</sub>	r <sub>270</sub>	p <sub>300</sub>
$\sigma_g$	246	240	241	242	241	243	243	242	242
$r_g$	0.90	0.94	0.97	0.98	0.99	1.00	1.00	1.00	1.00
$\sigma_p^2$	1459	1132	991	900	829	787	755	729	716
h)	0.21	0.24	0.26	0.28	0.30	0.31	0.32	0.33	0.34

Table 4. Variance in daughter average for  $\delta_x$  explained by breeding value for 305 day production (ebv<sub>305</sub>), breeding value for  $\beta_1$ (ebv<sub>B1</sub>), breeding value for  $\beta_2$ (ebv<sub>B2</sub>) and combinations of those variables.

	ebv <sub>305</sub>	ebv <sub>B1</sub>	ebv <sub>B2</sub>	ebv <sub>305</sub> + ebv <sub>B1</sub>	ebv <sub>B1</sub> + ebv <sub>B2</sub>	ebv <sub>305</sub> + ebv <sub>B1</sub> + ebv <sub>B2</sub>
$\delta_{60}$	1.8	19.4	6.7	19.7	33.6	33.8
$\delta_{90}$	2.5	26.0	0.7	26.4	30.8	30.7
$\delta_{120}$	1.6	25.7	0.0	25.8	26.6	26.6
$\delta_{150}$	1.2	24.4	0.8	24.4	24.4	24.4
$\delta_{180}$	1.8	21.5	1.3	21.8	21.5	21.8
$\delta_{210}$	0.7	14.6	2.7	14.6	15.1	15.3
$\delta_{240}$	0.7	6.3	1.6	6.4	6.7	7.0
$\delta_{270}$	0.2	5.4	1.7	5.3	5.9	5.9

Table 5. Variance in daughter average for  $\delta_x$  explained by breeding value for 305 day production (ebv<sub>305</sub>), breeding value for  $\beta_1$ (ebv<sub>B1</sub>), breeding value for  $\beta_2$ (ebv<sub>B2</sub>) and combinations of those variables, based on data from sires that have breeding values with reliabilities above 95%.

	ebv <sub>305</sub>	ebv <sub>B1</sub>	ebv <sub>B2</sub>	ebv <sub>305</sub> + ebv <sub>B1</sub>	ebv <sub>B1</sub> + ebv <sub>B2</sub>	ebv <sub>305</sub> + ebv <sub>B1</sub> + ebv <sub>B2</sub>
$\delta_{60}$	0.4	42.9	16.1	42.6	70.4	70.4
$\delta_{90}$	0.9	52.3	5.2	52.3	65.7	65.5
$\delta_{120}$	0.4	55.7	0.0	55.5	57.7	57.4
$\delta_{150}$	0.7	55.2	0.0	52.4	53.2	53.0
$\delta_{180}$	1.4	53.9	0.0	54.1	53.8	54.0
$\delta_{210}$	1.8	34.7	3.8	35.2	35.5	36.5
$\delta_{240}$	0.0	35.1	3.3	34.7	35.7	35.3
$\delta_{270}$	0.0	18.8	0.0	18.2	18.7	18.1

Table 6. Genetic covariance ( $\sigma_g$ ) between zx and  $p_{300}$ , genetic correlation ( $r_g$ ) between zx and  $p_{300}$ , and heritability (h) and phenotypic variance ( $\sigma_p^2$ ) of zx for protein yield.

	z <sub>60</sub>	z <sub>90</sub>	z <sub>120</sub>	z <sub>150</sub>	z <sub>180</sub>	z <sub>210</sub>	z <sub>240</sub>	z <sub>270</sub>	p <sub>300</sub>
$\sigma_g$	264	256	253	253	249	248	246	243	242
$r_g$	0.95	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\sigma_p^2$	1461	1134	994	903	831	789	757	730	717
h)	0.22	0.24	0.26	0.29	0.31	0.32	0.33	0.34	0.34