

# Statistical Indicators

## E-42

### Breeding value estimation for milk protein traits

#### ▪ Introduction

What really counts in the livestock industry is to breed a subsequent generation of cattle with a more profitable production by selection and mating. In this process the selection of milk protein traits is of increasing importance. These traits concern the production of specific milk proteins with economic value, such as casein (cheese making) and  $\beta$ -lactoglobulin (neonatal nutrition). The main aim of selection for protein traits is to improve the composition of milk. Breeding values for milk proteins are a tool to identify animals that will pass on their superior qualities to the next generation. In this part we will explain the background of the calculation of the breeding values for milk protein traits.

#### ▪ Overview

There are five milk proteins for which test day records are analysed. Additionally the traits milk production and total protein production are analysed. Hence we have 7 main production traits:

1. Milk production
2. Total protein
3. Casein
4.  $\beta$ -Lacto globulin
5.  $\alpha$ -Lactalbumin
6. Immunoglobulin G
7. Bovine serum albumin

An overview of the content of these proteins in the total protein fraction of milk is given in Table 1. The coefficient of variation (CV) is an indicator of how efficient selection will improve a trait. A trait with a high CV will improve relatively more after a round of selection than a trait with a low CV.

In the NL-TDM cows and bulls get breeding value for the daily production for each day between day 5 and day 420 in lactations 1, 2 and 3 for milk proteins. The statistical model used for analysing milk protein traits is identical to the test day model used for milk production. For more in

**Table 1** Phenotypic mean and standard deviation, coefficient of variation and fraction in total protein for milk proteins in the milk protein breeding value estimation (N = 8.512.985 test day records)

Protein	Abr.	mean (g/1000g milk)	Stdev (g/1000g milk)	CV (mean/stdev)	Fraction (% total protein)
Total milk protein	prt	35.5	4.26	0.12	100
casein	cas	27.7	3.74	0.13	76.9
$\beta$ -Lacto globulin	$\beta$ Lg	4.11	0.86	0.20	11.0
$\alpha$ -Lactalbumine	$\alpha$ La	1.13	0.11	0.09	3.4
Immuno globulin G	IgG	0.41	0.12	0.29	1.1
Bovine serum albumin	BSA	0.29	0.07	0.24	0.8

depth information on this model we refer to chapter E-7 'Breeding value estimation of milk production traits with test-day model'.

## ▪ **Basic principles for the correct calculation of the breeding value**

The following three items are essential for the calculation of the correct and true breeding values:

1. reliable milk protein production data;
2. reliable pedigree data;
3. the correct statistic model.

These three items will be discussed below:

### **Reliable milk production data**

Daily productions must meet the following requirements for use in the NL-TDM:

1. the cow must be herdbook registered;
2. only official daily productions are to be included, these can also be owner sampled production records;
3. only daily productions in the first three lactations of a cow are to be included;
4. only daily productions from day 5 on up to day 420 after calving are to be included;
5. cow must be housed in a known herd on test day;
6. the minimum age at calving is 640 days;
7. the sire of the cow must be known;
8. protein content below 10%;
9. calving interval of the current lactation must be at least 215 days;
10. no daily yields of a lactation must have the status "unreliable";
11. difference between realized and predicted yields must be within limits.

Requirement 11 is introduced to exclude records from the breeding value estimation with a large likelihood of errors in the data. The requirement is implemented as described below.

For each day in milk in each lactation of each cow a predicted yield is calculated with a Wilmink's lactation curve. For a Wilmink curve it is necessary to have at least 3 test days in a lactation. Difference between realized and predicted yield is calculated. The standard deviation of this difference is stored for each calving year, lactation number and day in milk. Realized yields are not included if the yield is larger than the predicted yield plus 6 times the standard deviation of the difference between predicted and realized yield. Such yields have a high likelihood of data errors.

There are no requirements on the number of test days in a single lactation, lactations with one test day can also be included. Also, it is not necessary for a cow to have test days of the heifer lactation included. Test days from cows without the heifer lactation (e.g. imported cows, cows that calved too early as a heifer) are used in the NL-TDM as well ensuring that these cows get a breeding value as well.

### **Reliable pedigree data**

An important aspect of the animal model is that the family relations between animals are included in the calculation of the breeding values. Both the previous generation of the animal (parents) as well as the generation after the animal (offspring) is taken into account.

The full official pedigree of an animal is included in the NL-TDM. In the end each pedigree will end up with parents, which are unknown. These unknown parents will be grouped and the average hereditary qualities of this group will become the predictor for their offspring.

The groups will be made up in a way that the parents with expected equal hereditary qualities will be placed in the same group (refer to the section 'Phantom groups'). After checking all ancestors then all animals will be examined whether they are directly useful for the animal model, in other words, whether an animal really adds information to the system. Animals, which do not add information to the system, must meet all the requirements below:

1. no daily productions in the NL-TDM;
2. no known parents;
3. only 1 offspring.

These animals will be considered as unknown in the NL-TDM and will be replaced by a phantom group.

Because the livestock industry does not only select animals within their own breed, but also makes use of animals from other breeds, there are many genetic (family) relations between various breeds. For example, the use of Holstein Friesians in the Dutch Friesian breed and in the MRY population. The many genetic relations between breeds makes it possible to calculate the breeding values for various breeds simultaneously, which is performed in the NL-TDM.

### Phantom groups

When a parent is unknown a so-called "phantom parent" replaces this parent. Phantom parents can be grouped in phantom groups or genetic groups. Such a phantom group provides a kind of substitute genetic value of an unknown parent. The genetic value of such a group is determined by all animals, which are connected to this specific group. So, it is essential to put animals with most likely the same genetic potential all together in one group. Circumstances, which can cause a difference in genetic potential and because of this are essential for the definition of the phantom groups, are:

#### *1. Selection path*

The selection path is subdivided into the following six groups ( I through VI):

- for cattle with one unknown parent:
  - I. cows with an unknown sire
  - II. cows with an unknown dam
  - III. bulls with an unknown sire
  - IV. bulls with an unknown dam
- for cattle with two unknown parents:
  - V. cows with two unknown parents
  - VI. bulls with two unknown parents

#### *2. Breed*

A separate group will be made up for each breed. In practice this means that there are approximately 30 various breed groups.

In case of a crossbred animal, this animal will be classified in the breed with the highest share of genes. The following two situations occur in case both blood breed shares are equal:

- for 50%/50% FH/HF-breeds and 50%/50% MRY/HF-breeds, there are two separate "breed" groups; these animals will not be linked to one of the parent breeds;
- for the remaining 50%/50%-breeds: the animal will be classified into one of the parent breeds based on the breeding goal of the crossbred animal. There are three breeding goals: milk, dual purpose and beef. The breeding goal determines the major breed; beef before dual purpose before milk. If both breeds have the same breeding goal, the first breed is considered the major breed.

#### *3. Country of origin*

Each country has a characteristic genetic level for its breeds, this way groups can be classified in the country of origin.

#### 4. Year of birth

Each year of birth forms a group. Cattle born before 1950 are linked to the year of birth 1950.

A phantom group is made up by each combination of selection path, breed and year of birth. A maximum of  $k \cdot l \cdot m \cdot n$  phantom groups can be formed by “k” (=6) groups for selection paths, “l” groups for breeds, “m” groups for countries and “n” groups for years of birth.

The minimum number of animals (parents) per phantom group is 40. Phantom groups will be combined in case the number of 40 is not reached. Only years of birth within the selection path \*breed\*-country combination will be put together in one group. A combination per year of birth will be made until the number of 40 is reached or when at the most 10 years of birth are combined together. In case the number of 20 animals is not reached after the combination of 10 years of birth, then a combination of the phantom group will be made with all other selection paths, countries and years of birth within the breed. When a breed is so small and contains less than 20 animals, then the phantom group will be added to the remaining group. This remaining group will be made up of various phantom groups, which do not contain enough animals to be able to make a reliable estimation. There are three remaining groups, by production goals: milk breed, dual purpose and meat.

The influence of the phantom group estimation on breeding values declines more and more as the phantom group is generations removed from the specific animal and/or as the cow or bull gets more information by offspring and/or characteristic lactation(s).

#### The statistic model

The calculation of the breeding values uses the BLUP-technique (Best Linear Unbiased Prediction) in an animal model. The characteristic feature of this animal model is that breeding values of cows and bulls are estimated simultaneously and all genetic relations are taken into account. The statistic model for analysing the daily production is as follows:

$$y_{ijklmnopqrst} = \text{HTD}_i + \text{DIM} \times \text{PAR}_j + \text{PAYS}_{-L_k} + \text{PAYS}_{-C_l} + \text{PREG} \times \text{PAR}_m + \text{HET} \times \text{PAR}_n + \text{REC} \times \text{PAR}_o + \text{DRYxcDIM}_p + \text{animal}_q + \text{perm}_r + \text{herd}_s + \text{res}_{ijklmnopqrst}$$

in which:

$y_{ijklmnopqrst}$	:	production of trait ;
$\text{HTD}_i$	:	herd-test date $j$ ;
$\text{DIMxPAR}_j$	:	parity x day in milk $k$ ;
$\text{PAYS}_{-L_k}$	:	parity x age at calving x year x season of calving $l$ ;
$\text{PAYS}_{-C_l}$	:	parity x age at calving x year x season of calving x lactation stage $m$ ;
$\text{PREGxPAR}_m$	:	parity x stadium gestation $n$ ;
$\text{HETxPAR}_n$	:	parity x class of heterosis $o$ ;
$\text{RECxPAR}_o$	:	parity x class of recombination $p$ ;
$\text{DRYxcDIM}_p$	:	class of number of days dry x lactation stadium $q$ ;
$\text{animal}_q$	:	additive genetic effect (or breeding value) of cow $r$ ;
$\text{perm}_r$	:	permanent environmental effect of cow $r$ ;
$\text{herd}_s$	:	herd x year of test $s$ ;
$\text{res}_{ijklmnopqrst}$	:	residual term of $y_{ijklmnopqrst}$ , which is not explained by the model.

The effects *animal*, *perm* and *herd* are random effects fitted on a 420-day lactation curve. The *animal* effect is additionally uses a full pedigree to account for relations between animals.

*Lactation curve* The average lactation curve of the entire population is modelled with the effect of DIMxPAR. This effect represents the average production on each day between (and including) day 5 and day 420 in parity 1, 2 and 3. PAYS\_L models the differences in production level between cows, which differ in age and year-season of calving from one another. This effect cannot model the difference in the course of the production curve; the curves differ only in level from one another. The effect PAYS\_L contains many classes for different ages at calving (for example 26 or 30 months), because it is known that these cows indeed differ in production level from one another, but not in the course of the production curve. With the effect PAYS\_C differences in the course of the production curve are modelled. The effect PAYS\_C contains many classes for different seasons of calving (for example calving in February or July), because it is known that these cows differ in the course of the production curve from one another.

*HTD* The herd level on the test day is modelled with the herd test day effect. With this effect daily productions of cows can be compared with other cows in the herd, which have produced on the same test day. No distinction will be made between cows in different parities. They are supposed to have had the same management. Cows of the same herd from different breeds are also compared directly to one another. The possible genetic differences between breeds will be compensated by the additive genetic effect. Herd test days with little observations will not be combined. If there are no other cows present in the same herd then an observation will not contribute in the breeding value of the cow.

In the lactation model each lactation had to be assigned to one specific herd. This is difficult for cows, which move to other herd during the lactation. In the test day model cows are compared with other cows in the same herd on the same test day. Before moving to another herd they will be compared with other cows of the same first herd, after moving they will be compared with other cows of that specific herd they have moved to.

*Stage of gestation* The effect of gestation on the milk protein production will be corrected using PREGxPAR. It is known that cows in calf have less production, especially during the last stadium of the gestation, just before the dry condition.

*Days dry x stage of lactation* The days dry x stage of lactation effect can correct the negative effect of a short dry period on the production in the following lactation. The length of the dry period is calculated as the number of days between the last sampling from the previous lactation and the following calving date. Research shows that cows with very short periods of dry condition (for example <40 days) have a lower production in the following lactation, especially in the beginning of this lactation. It is likely that most of these cows have a short dry period, due to an abortion. Cows with an average or long period of dry condition do not differ in production. Because an abortion can be regarded as a non-genetic effect, it is justified to correct this in the breeding value estimation for production. Because of the interaction with the lactation stadium it is possible that corrections can be different for various lactation stadiums.

*Heterosis-effect and recombination effect* play a part in the cross breeding of breeds. These are genetic effects, which will not be passed on to offspring. Research shows that these effects must be corrected.

The size of the heterosis is defined as the difference in level of the trait in the cross breeding with the average of the parent breeds. Recombination is the loss of, in most cases, the positive effect of heterosis and occurs when the crossbred product, obtained earlier, is crossed back with one of the parent breeds.

The same correction factors are used for all cross breeding combinations. The corrections are different for lactation 1, 2 and 3, but the same correction is performed on each day within the lactation.

*Animal effects* There are two kinds of animal-related effects in the model: the additive genetic effect and the permanent environmental effect. The additive genetic effect models the genetic hereditary qualities of an animal on each day in lactation 1, 2 and 3. In this context additive means that these hereditary qualities can be passed on to offspring.

The permanent environmental effect models the part of the production graph of a cow, which cannot be explained by the fixed effects or the additive genetic effect. For example, if a cow has had a disease during her breeding, then this can have a negative influence on her future production. However, the fixed effects do not explain this case and it is also not a genetic effect, because this cannot be passed on to an offspring. The lower production can partly be compensated by the permanent environmental effect and does not need to result in a lower breeding value. To which extent the breeding value becomes lower depends on the heritability (in case of a high heritability the individual production of a cow has more influence on her breeding value) and the available information on the members of the family (if a cow does not have any offspring then, to a large extent, the lower production will be in the permanent environment and not in the breeding value).

The additive genetic effect and the permanent environmental effect are random effects, which means that the amount of available information on an animal is taken into account. If there is little information on an animal (not many offspring and too little test day productions), the additive genetic effect will not differ much from the parent average. The permanent environmental effect will be close to zero.

*Herd specific production curves* The *herd* effect results in curves for every herd x year of test. These curves are calculated from the production data of all cows in a herd within one calendar year. The herd specific curves are used to correct for the effect of herd management of production circumstances on the production curve. For example, it is likely that a herd with a flat feeding systems may have a ration that is relatively poor for fresh cows and rich for cows later in lactation. These effects are not genetic, so they must be accounted for. The herd specific curves are included as random effects to account for variations in herd sizes. The herd-test-date effect in the NL-TDM models the mean production of the herd, whereas the herd curves model the shape of the production curve.

### Random regression

The effects *animal*, *perm* and *herd* are random regression effects. This means that the *animal*, *perm* and *herd* effect are not modelled with a single number but with a curve that ranges from day 5 to 420 and for lactation 1, 2 and 3.

The additive genetic effect and the permanent environmental effect are modelled by means of a random regression function. In this case the regression function is a mathematical function, which describes the additive genetic effect or the permanent environmental effect on the daily production on each day between day 5 and day 420 in lactations 1, 2 and 3. These effects need not to be the same on each day in each lactation. So, it is possible that, for example, animals are genetically superior in the beginning of the lactation, but are less so in the end of the lactation. A regression function for a single lactation can be as follows:

$$y_r(DIM) = z_0 \times a_{r0} + z_1 \times a_{r1} + z_2 \times a_{r2} + z_3 \times a_{r3}$$

in which:

- $y_r(DIM)$  : additive genetic effect or permanent environmental effect of an animal  $r$  for daily production on  $DIM$ ,
- $z_i$  : regression coefficient  $i$ . The regression coefficients are dependent on  $DIM$ .
- $a_{ri}$  : solution of an animal  $r$  for the regression coefficient  $i$ ;
- $DIM$  : days in milk

An example of a regression function for one lactation:  $z_0 = 1$ ,  $z_1 = x$ ,  $z_2 = x^2$ ,  $z_3 = x^3$ , in which  $x = 2(DIM-5)/(420-5) - 1$  and DIM is a number between 5 and 420. This ensures  $x$  is a number between  $-1$  and  $1$ , which is a requirement for the mathematical function used to model individual curves.

The additive genetic effect or the permanent environmental effect on DIM=46 can be calculated as follows:

$$x = 2*(46 - 5)/(420 - 5) - 1 = -0.80$$

$$\begin{aligned} y_r(38) &= a_{r0}*(1.00) + a_{r1}*(-0.80) + a_{r2}*(-0.80)^2 + a_{r3}*(-0.80)^3 \\ &= a_{r0}*(1.00) + a_{r1}*(-0.80) + a_{r2}*(0.64) + a_{r3}*(-0.512) \end{aligned}$$

Suppose that the solutions of an animal  $r$  for casein production are equal to  $a_{r0}=+450$ ,  $a_{r1}=-150$ ,  $a_{r2}=-30$ , and  $a_{r3}=15$ . Then  $y_r(46)$  can be calculated:

$$\begin{aligned} y_r(46) &= 450*(1.00) + (-150)*(-0.800) + (-30)*(0.640) + 15*(-0.512) \\ &= 450 + 120 + (-19.2) + (-7.7) \\ &\approx 543 \text{ g casein} \end{aligned}$$

More information about random regression can be found in chapter E-7 'Breeding value estimation of milk production traits with test-day model'.

Note that the animal solutions  $r$  can be used to predict the daily production for each day in a lactation from day 5 to day 420. For instance, the above solution gives the following prediction for DIM=286:

$$\begin{aligned} x &= 0.354 \\ y_r(286) &= a_{r0}*(1.00) + a_{r1}*(0.354) + a_{r2}*(0.354)^2 + a_{r3}*(0.354)^3 \\ &= 450 + (-150)*0.354 + (-30)*0.125 + 15*0.044 \\ &\approx 394 \text{ g casein} \end{aligned}$$

Hence, once we have animal solutions for the regression coefficients, we can calculate the genetic merit at each day in the lactation. The solutions for the additive genetic effect are the breeding values belonging to the regression coefficients. Animals with a high breeding value for the first regression coefficient will have a high  $y_r$ , regardless the value of DIM. So, these animals will also have a high breeding value for 305-daily production. The second regression coefficient is low at the beginning of the lactation ( $z_1=-1$ ) and high at the end ( $z_1=1$ ). This means that animals which have a high breeding value for the second regression coefficient will have a high breeding value at the end of the lactation and a low breeding value at the beginning of the lactation.

In the actual NL-TDM there are 5 regression coefficients for the additive genetic effect and 5 for the permanent environmental effect for each protein trait and each lactation in the NL-TDM. This means that a single genetic curve (one trait in one lactation) consists of 5 breeding values. In total there are 7 traits and 3 lactations. A 305-day cumulative breeding value can then be calculated by modelling the curve using the 5 animal effects solutions and summing breeding values per DIM from day 5 to day 305.

#### *Published breeding values*

The milk and protein production breeding values in the milk protein breeding value estimation are not published. They do not have official status, since they are estimated only for the subset of animals that have a record or offspring with records for milk proteins. However, these breeding values are used to derive breeding values for content traits.

Breeding values, which are derived in the NL-TDM for milk proteins, are:

1. 305-day production casein in lactation 1
2. 305-day production casein in lactation 2
3. 305-day production casein in lactation 3
4. 305-day production casein (overall)
5. 305-day production  $\beta$ -Lacto globulin in lactation 1
6. 305-day production  $\beta$ -Lacto globulin in lactation 2
7. 305-day production  $\beta$ -Lacto globulin in lactation 3
8. 305-day production  $\beta$ -Lacto globulin (overall)
9. 305-day production  $\alpha$ -Lactalbumin in lactation 1
10. 305-day production  $\alpha$ -Lactalbumin in lactation 2
11. 305-day production  $\alpha$ -Lactalbumin in lactation 3
12. 305-day production  $\alpha$ -Lactalbumin (overall)
13. 305-day production Immunoglobulin G in lactation 1
14. 305-day production Immunoglobulin G in lactation 2
15. 305-day production Immunoglobulin G in lactation 3
16. 305-day production Immunoglobulin G (overall)
17. 305-day production Bovine serum albumin in lactation 1
18. 305-day production Bovine serum albumin in lactation 2
19. 305-day production Bovine serum albumin in lactation 3
20. 305-day production Bovine serum albumin (overall)

The overall breeding values for 305-daily production are calculated as follows:

$$BV_{total} = 0.41 \times BV_1 + 0.33 \times BV_2 + 0.26 \times BV_3$$

in which:

$BV_i$  : breeding value for 305-daily production in lactation  $i$ .

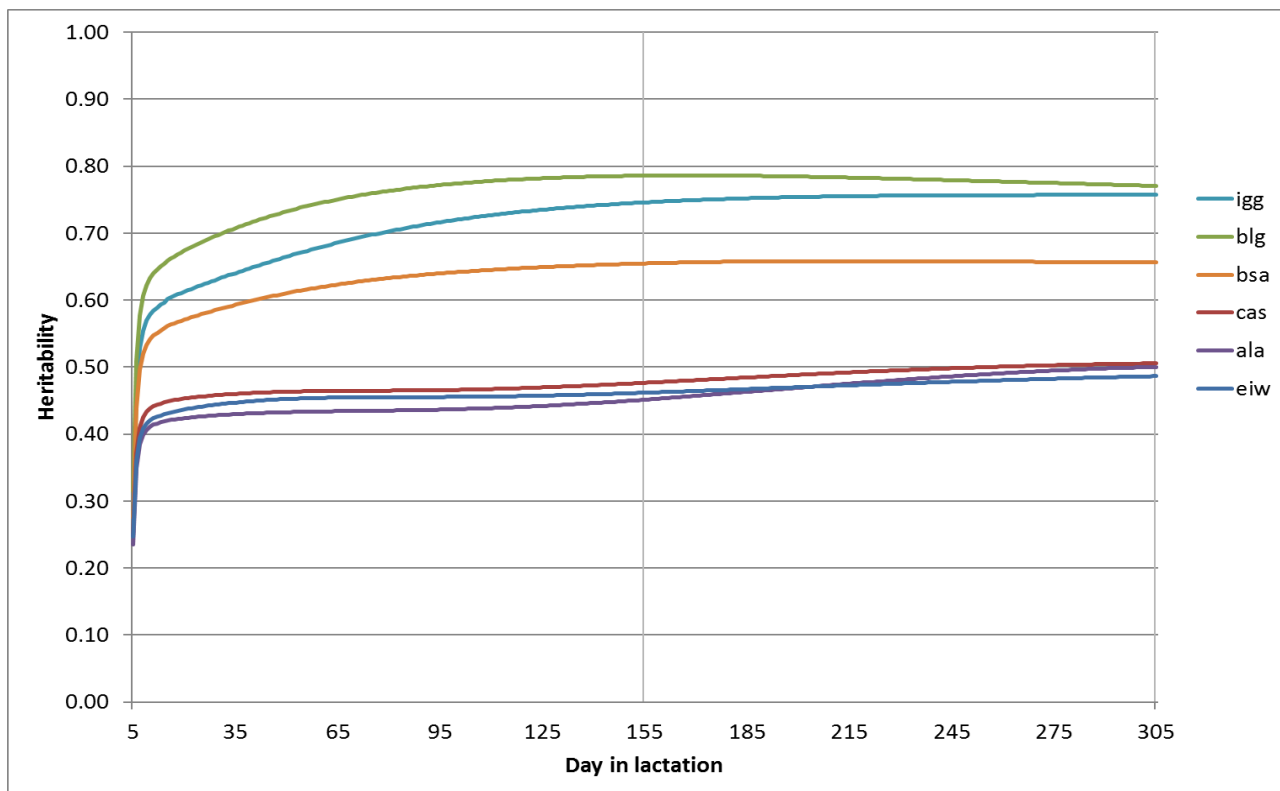
The factors (0.41, 0.33 en 0.26) are primarily based on the distribution of 1<sup>st</sup>, 2<sup>nd</sup> and 3<sup>rd</sup> heifer cows in the Netherlands. Furthermore, the income from lactation 2 and 3 are discounted for the loss of interest due to the fact that a longer waiting period for this income is needed. Beside this an additional weight is inserted in lactation 2 and especially in lactation 3, because these lactations are stronger related with the production in lactation 4 and higher.

### Genetic parameters

Parameters are required for the breeding value estimation of effects, which are regarded as random effects (the additive genetic effect, permanent environmental effect, the herd curves and residual term). These are, for example, the genetic variances, heritabilities and the correlation between various traits. These parameters are estimated on a part of all national data.

Each day between 5 and 420 in lactations 1, 2 and 3 is regarded as a different trait in the NL-TDM. This means that an animal, which is genetically good at the beginning of the lactation, does not have to be good at the end of the lactation or an animal, which is good in lactation 1, does not have to be good in lactation 3. Nevertheless, breeding values in various days and lactations are correlated to one another and the closer the days are to one another the higher the genetic correlation is. Each milk protein is analysed in a separate run of the NL-TDM. This means that correlations between milk proteins are not accounted for. Figure 1 shows the heritability for each day for the yield breeding values of the five milk proteins. The average heritability for daily yield is





**Figure 1.** Heritability for Immunoglobulin G (**igg**),  $\beta$ -Lactoglobulin (**blg**), Bovine Serum Albumin (**bsa**), Casein (**cas**) and  $\alpha$ -Lactalbumin (**ala**) during the lactation.

48% for casein and  $\alpha$ -lactalbumin production, 64% for bovine serum albumin production and 74% for  $\beta$ -lactoglobulin and immunoglobulin G production.

Genetic correlations between traits indicate how similar the breeding values for the specific traits are. This also indicates how well the breeding value of the one trait can be predicted from the breeding value of the other trait. Table 2 shows the heritabilities (diagonal) and genetic correlations for 305-day production for the milk proteins in various lactations and also the correlations with the overall breeding value.

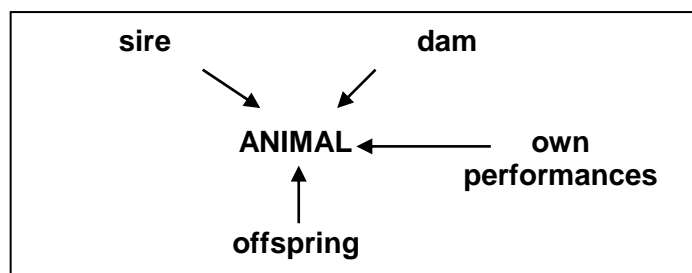
### Reliability

Breeding values are estimations of the hereditary qualities. The word “estimation” indicates that there is a certain inaccuracy resting on a breeding value. The reliability of a breeding value indicates the dimension of the difference between the estimated breeding value and the actual hereditary qualities.

The reliability depends on the quantity of the available information of an animal. There are three sources of information:

1. own performance
2. offspring
3. parents

Information on the production of (half) sisters, grandparents, etc. is inserted through the parents, information of granddaughters etc. is inserted through the offspring.



**Table 2.** Heritabilities (diagonal) and genetic correlations between lactations 1, 2 and 3 and the overall breeding value for milk protein traits.

	Parity 1	Parity 2	Parity 3	Overall
<b>Casein</b>				
Parity 1	<b>0.30</b>	0.60	0.52	0.84
Parity 2		<b>0.34</b>	0.58	0.87
Parity 3			<b>0.39</b>	0.82
Overall				<b>0.51</b>
<b>bLg</b>				
Parity 1	<b>0.52</b>	0.80	0.77	0.93
Parity 2		<b>0.61</b>	0.77	0.94
Parity 3			<b>0.60</b>	0.91
Overall				<b>0.77</b>
<b>aLa</b>				
Parity 1	<b>0.32</b>	0.58	0.54	0.84
Parity 2		<b>0.37</b>	0.62	0.87
Parity 3			<b>0.39</b>	0.83
Overall				<b>0.50</b>
<b>IgG</b>				
Parity 1	<b>0.51</b>	0.80	0.73	0.92
Parity 2		<b>0.59</b>	0.78	0.94
Parity 3			<b>0.60</b>	0.90
Overall				<b>0.76</b>
<b>BSA</b>				
Parity 1	<b>0.42</b>	0.73	0.64	0.89
Parity 2		<b>0.49</b>	0.69	0.92
Parity 3			<b>0.48</b>	0.86
Overall				<b>0.66</b>

Because each day in each lactation is regarded as genetically different traits, the breeding values for each day have their own reliability. The breeding value for 305-day production in lactation 1 can have a different reliability than the breeding value for 305-day production in lactation 3. If a bull only has 1<sup>st</sup> parity daughters, then the breeding value for lactation 3 will have a lower reliability than the breeding value for lactation 1. The breeding value for lactation 3 will then, on average, deviate less from the parent average than the breeding value for lactation 1. This does not mean that the breeding value for lactation 3 is exactly the same as the parent average. The 1<sup>st</sup> parity daughters of a bull already show whether the bull is better than his parent average in lactation 1. This also provides information on the difference with the parent average in lactation 3, because the breeding values of lactations 1 and 3 are correlated.

#### Heterogeneity of variance

The differences in variation of the daily production on various herd test days (HTD) and the differences in variation of the daily production of cows of different ages are taken into account in the breeding value estimation. It is clear that, compared to other herds, some herds have more variation in their daily production. This can be caused by the differences in ages and patterns of calving between herds. For example, it is likely that on a herd test day, where only heifers at the end of their lactation are milked, the variation in daily production is lower than on a herd test day, where both heifers at the end of their lactation as well as fresh 3<sup>rd</sup> parity cows are milked. Another source for the differences in variation between herd test days can lie in the hereditary qualities of the cows. On a herd test day with only MRY-cows the variation will be less than on a herd test day with both HF-cows as well as Belgian Blue cows. These two sources of variation have no influence in the correctness of the breeding value estimation.

A third source for variation is the difference in management between herds, for example, forage stock, housing accommodation and the breeding of young cattle. These differences may be the reason that genetic differences are expressed more on one herd than on the other. When this source is not taken into account in the breeding value estimation, then the best cows of the herd with a high variation will have higher breeding values than cows of a herd with a lower variation. The reason for this is that the best cows on a herd with a high variation deviate more from their herd mates.

The correction for heterogeneity of variance in the NL-TDM is performed by multiplying the permanent environmental effect and the residual part of a daily production (this is the production minus the solutions of all fixed effects and the additive genetic effect) by a correction factor. For details see chapter E-7.

## ▪ Base definitions

Breeding values for the milk protein production of bulls and cows are published based on the 2015-base. The average breeding value (for each trait) of cows in the base population is set to 0. Breeding values of all animals are expressed this way in relation to their base. Cows born in 2010 determine the base of 2015. There are four different bases: Milk goal Black, Milk goal Red, Dual purpose and Belgian Blue. The definitions of these bases are as follows:

### *Milk goal Black*

Herdbook-registered cows born in 2010 with at least 87.5% HF-genes and up to 12.5% FH-genes and hair colour black pied, with at least one observation in the genetic evaluation.

### *Milk goal Red*

Herdbook-registered cows born in 2010 with at least 87.5% HF-genes and up to 12.5% MRY-genes and hair colour red pied, with at least one observation in the genetic evaluation.

### *Dual purpose*

Herdbook-registered cows born in 2010 with at least 75% MRY-genes and 25% or less HF genes, with at least one observation in the genetic evaluation.

### *Belgian Blue*

Herdbook-registered cows born in 2010 with at least 87.5% Belgian Blue-genes, with at least one observation in the genetic evaluation.

### *Heterogeneity of variance*

The base for the correction factor for heterogeneity of variance is determined by HF/FH-cows born in 2000. This basis is applied to all breeds.

## ▪ Calculation of % casein and % casein in protein

The breeding values for casein content are calculated from the breeding values for kg milk, kg protein and kg casein using phenotypic productions, as shown below:

$$BV_{\%casein} = \frac{BV_{kgcasein} * 100 - BV_{kgmilk} * P_{\%casein}}{BV_{kgmilk} + P_{kgmilk}}$$

$$BV_{\%cas\_in\_prot} = \frac{BV_{kgcasein} * 100 - BV_{kgprotein} * P_{\%cas\_in\_prot}}{BV_{kgprotein} + P_{kgprotein}}$$

in which:

P = average production of the base animals for the given traits;

BV = breeding value: kg milk, kg protein or kg casein, % casein, % protein or % cas\_in\_prot.

The breeding values for these content traits can be different for each day in each lactation. The average lactation production of the basic cows is used in the breeding values for the percentages on 305-day-level. The average lactation production is calculated as the sum of the average daily productions between day 5 and day 305 of the base cows. The average lactation productions  $P_{\text{kgm}}$ ,  $P_{\% \text{protein}}$  and  $P_{\% \text{cas\_in\_prot}}$  for the Black, Red, Dual purpose and Belgian Blue base are shown in Table 3. Overall % breeding values are calculated from the overall mean lactation productions, where the latter is the weighted average of the lactation productions in lactations 1, 2 and 3 using the same factors as in the calculation of the overall breeding value: 0.41, 0.33 and 0.26.

## ▪ Publication rules

Breeding values for milk protein production traits for AI bulls will be published when a bull has at least 15 daughters on at least 10 herds that have passed day 120 of lactation 1. Bulls are considered AI bulls when they have an AI code and an owner who is not registered as a farmer. Non AI bulls do not have to meet the publication rules mentioned above; they will be published as soon as they have a reliability of 35% or more and at least 10 daughters in milk recording.

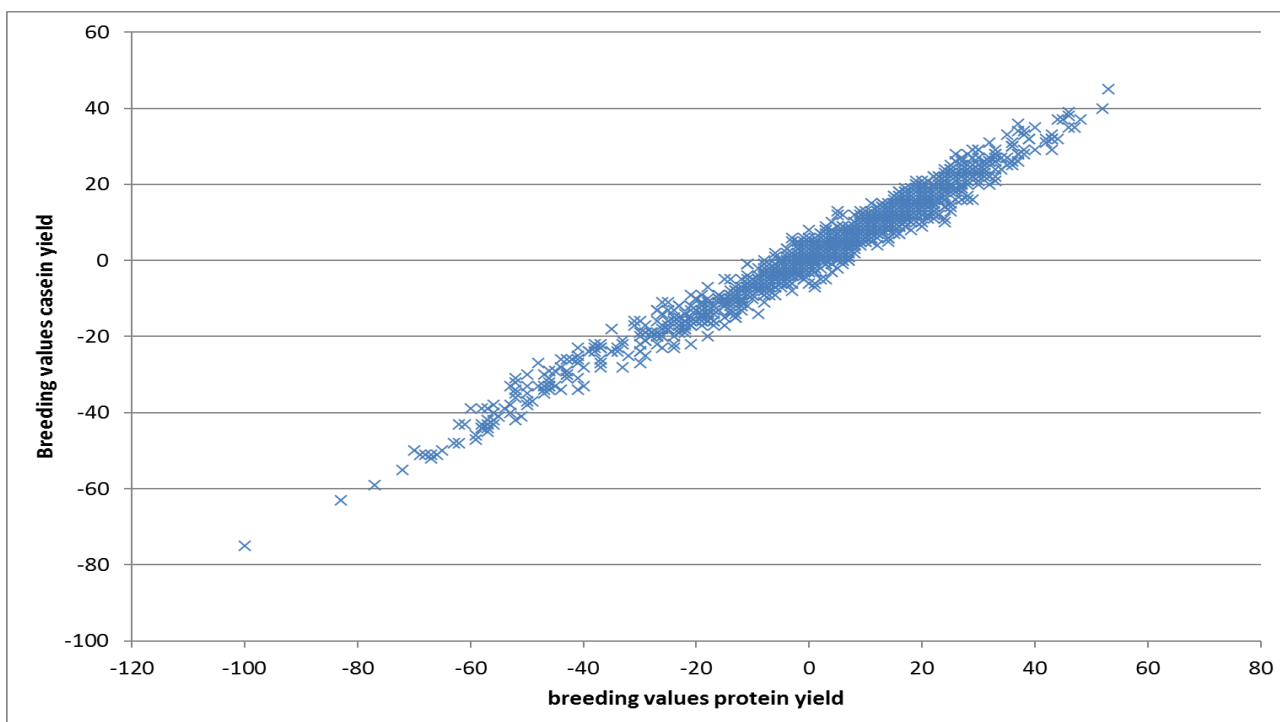
Breeding values for milk protein production traits for cows will be published when their sire has a publishable breeding value. There is no requirement with respect to length of lactation for cows. Daughters of bulls that do not have a publishable breeding value will get a breeding value themselves if they have passed day 220 of lactation 1. In this manner, cows will always get a published breeding value based on their own performance when they finish their first lactation.

**Table 3a.** Average productions used for calculation of content breeding values

		kg milk	kg protein	kg casein
Milk goal Black	Parity 1	7611	263	204
	Parity 2	8927	313	241
	Parity 3	9453	327	261
	<b>Overall</b>	<b>8524</b>	<b>296</b>	<b>228</b>
Milk goal Red	Parity 1	7051	251	196
	Parity 2	8238	299	232
	Parity 3	8808	314	243
	<b>Overall</b>	<b>7900</b>	<b>283</b>	<b>220</b>
Dual Purpose & Belgian Blue	Parity 1	5658	207	165
	Parity 2	6398	238	189
	Parity 3	6828	252	200
	<b>Overall</b>	<b>6206</b>	<b>229</b>	<b>182</b>

**Table 3b.** Average production content for protein, casein and casein in protein.

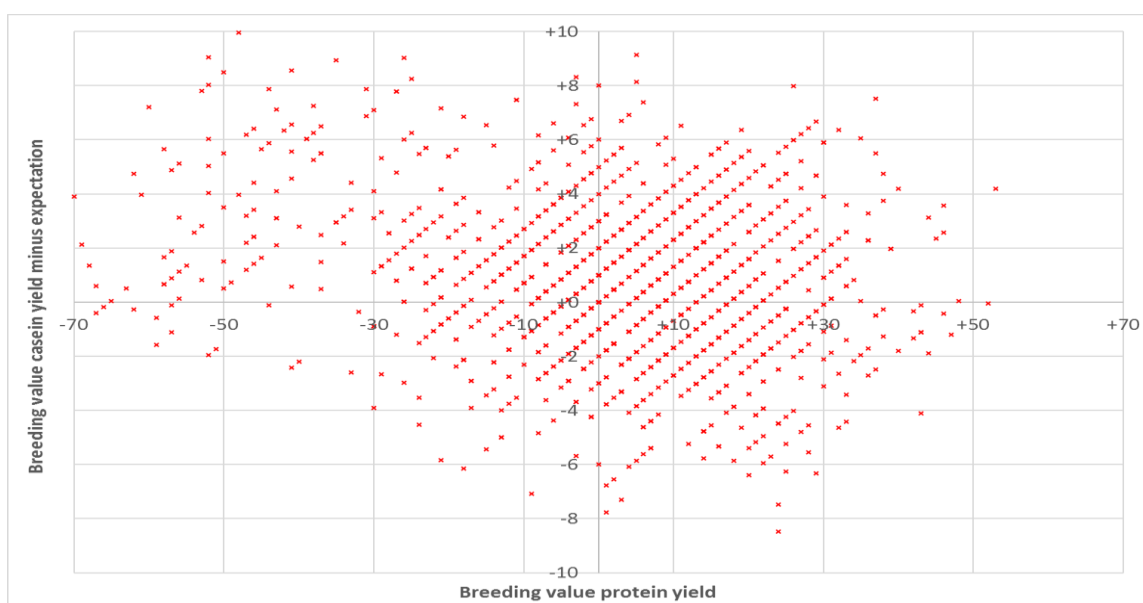
		% protein	% casein	% casein in protein
Milk goal Black	Parity 1	3.45	2.68	77.56
	Parity 2	3.51	2.70	76.93
	Parity 3	3.46	2.65	76.66
	<b>Overall</b>	<b>3.47</b>	<b>2.68</b>	<b>77.10</b>
Milk goal Red	Parity 1	3.56	2.78	78.17
	Parity 2	3.63	2.82	77.62
	Parity 3	3.57	2.76	77.36
	<b>Overall</b>	<b>3.59</b>	<b>2.79</b>	<b>77.72</b>
Dual Purpose & Belgian Blue	Parity 1	3.66	2.92	79.83
	Parity 2	3.72	2.96	79.41
	Parity 3	3.69	2.93	79.38
	<b>Overall</b>	<b>3.69</b>	<b>2.94</b>	<b>79.60</b>



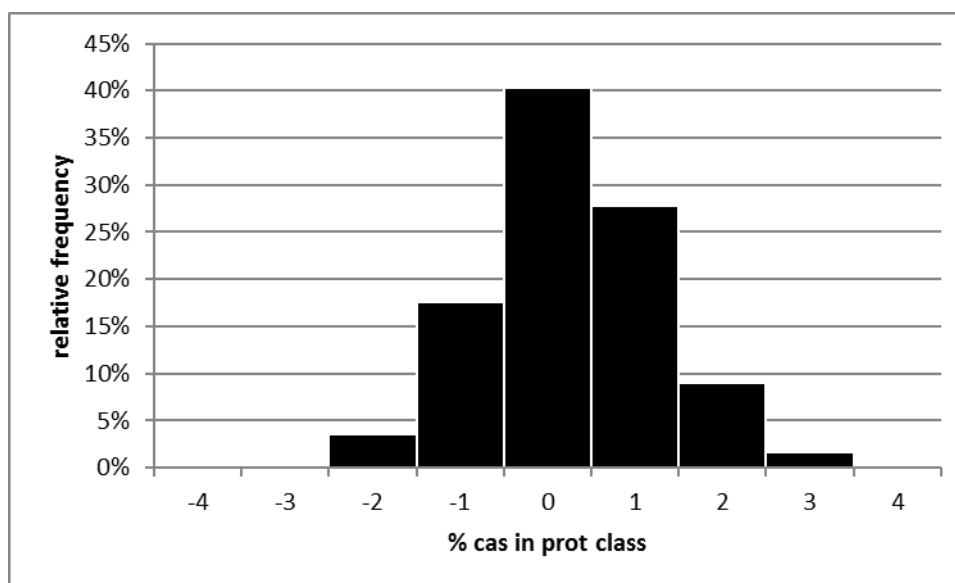
**Figure 2.** Scatterplot of breeding values casein yield versus protein yield (N = 1375 CRV bulls born between 2005 and 2010). Correlation between breeding values is 0.98.

### ▪ Casein breeding values yield and content

The breeding values for production of protein and casein are highly similar (see Figure 2). The correlation between protein and casein breeding values is 0.98. Nevertheless casein breeding values show some variation relative to protein breeding values (see Figure 3). For example: Bulls with a breeding value of +22 kg protein showed a range of breeding values for kg casein from +11 kg to +22 kg. This shows that despite the high correlation there is some room for improving casein production relative to protein production.



**Figure 3.** Scatterplot of breeding values casein yield minus expected breeding values casein yield ( $= 0.77 * \text{protein yield}$ ), showing the variation in breeding values casein yield for each value of protein yield, despite the high correlation (N = 1375 CRV bulls born between 2005 and 2010).



**Figure 4.** Bar graph of frequency per casein content in protein class (defined as the breeding values % CinP rounded to the nearest percentage point)

### Casein content in protein

The trait 'casein content in protein' is defined as the casein production expressed as percentage of total protein production, "%CinP". Earlier analysis has shown some variation in this trait, despite the high correlation between protein yield and casein yield breeding values. The standard deviation in %CinP breeding values is nearly 1%, which is about the expected value. Values range between +2.7 and -3.8 % casein in protein. Figure 4 displays the distribution of breeding values rounded to the nearest integer percent point.

### Effects of selection

Table 4 gives an overview of correlations between breeding values for protein and casein. Both the breeding values for yield and content showed a high correlation: 0.98 in both cases. This correlation holds even when breeding values are analyzed per birth year. Hence, selection on protein yield only will increase the casein yield. However, correlations between yield and content breeding values are moderately negative. This means that persistent selection on protein yield only will slowly decrease casein content in both milk and protein.

The trait casein content in protein shows no correlation with casein yield (-0.07) and a relatively strong correlation with %C (0.66). However, the correlation with protein production is negative (-0.25). This implies that selection for this trait will not increase casein yield, but will reduce protein production resulting in higher casein content.

**Table 4.** Correlation between breeding values of protein and casein traits

	Protein yield	Casein yield	%P	%C	%CinP
Casein yield	0.98				
% protein	-0.32	-0.24			
% casein	-0.34	-0.23	0.98		
% CinP	-0.25	-0.07	0.48	0.66	
Milk yield	0.89	0.84	-0.72	-0.72	-0.41

Casein content in protein (%CinP) is heritable ( $h^2 \sim 0.3$ ) and shows considerable variation (gen. std. 0.99% ). However, the increase in casein production per point %CinP is modest. Per whole percent %CinP the casein yield is expected to increase with around 1.5 kg (range 1.3 – 1.8 kg). This means that a bull with a +1.00 breeding value %CinP will produce daughters which will yield 228.5 kg casein instead of 227 kg. For a bull with breeding value +3.00, daughters are expected to produce on average 231.5 kg casein.

Selection for casein yield will somewhat decrease protein and casein content in milk. Given the low correlation between casein yield and % casein in protein, such selection will have no effect on casein content in protein.

In Summary:

- Selection for protein yield will lower casein content
- Selection for casein yield will increase casein production but lower the casein and protein content of the milk.
- Selection on casein content in protein will increase casein content in milk, but lower protein yield.
- Simultaneous selection on casein yield and casein content in protein will increase casein and protein yield as well as casein content.