



## **A Stochastic Simulation Study on Validation of an Approximate Multitrait Model Using Preadjusted Data for Prediction of Breeding Values**

J Lassen, Mk Sorensson, P Madsen, Vincent Ducrocq, . Station de Génétique Quantitative Et Appliquée, . Department of Genetics And Biotechnology, .  
Department of Large Animal Sciences

### **► To cite this version:**

J Lassen, Mk Sorensson, P Madsen, Vincent Ducrocq, . Station de Génétique Quantitative Et Appliquée, et al.. A Stochastic Simulation Study on Validation of an Approximate Multitrait Model Using Preadjusted Data for Prediction of Breeding Values. *Journal of Dairy Science*, 2007, 90, pp.3002-3011. hal-02661965

**HAL Id: hal-02661965**

**<https://hal.inrae.fr/hal-02661965>**

Submitted on 30 May 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## A Stochastic Simulation Study on Validation of an Approximate Multitrait Model Using Preadjusted Data for Prediction of Breeding Values

J. Lassen,\*†<sup>1</sup> M. K. Sørensen,\* P. Madsen,\* and V. Ducrocq‡

\*Department of Genetics and Biotechnology, Danish Institute of Agricultural Sciences, P.O. Box 50, DK-8830 Tjele, Denmark

†Department of Large Animal Sciences, The Royal Veterinary and Agricultural University, Ridebanevej 12, DK-1870 Frederiksberg C, Denmark

‡Station de Génétique Quantitative et Appliquée, Institut National de la Recherche Agronomique, 78352 Jouy-en-Josas, France

### ABSTRACT

Three different models for prediction of breeding values were compared in a stochastic simulation study of a dairy cattle population of 100,000 cows. The simulation was done in 2 steps. The first step involved 15 yr of selection using breeding values obtained in a univariate model for production and a trivariate model for mastitis occurrence, udder depth, and somatic cell score, in which production and mastitis occurrence were included in the breeding goal. This was done to create an initial population that had already been under selection. The second step consisted of 20 replicates of 4 different scenarios set up to make it possible to compare the different models. Two scenarios were based on univariate evaluations and one for udder health traits on trivariate evaluations, with 2 different breeding goals. In another scenario, an approximate multitrait model using preadjusted data in a 2-step procedure was used and in the last scenario, a complete linear multitrait model was carried out. Differences in genetic response in total merit over the last 15 yr of selection were compared and used to rank the models. The linear multitrait model gave the highest regression coefficient of true genetic values on year ( $3.073 \pm 0.069$  in economic units), and this was significantly better than for the approximate multitrait model ( $2.819 \pm 0.047$ ), which again was significantly better than for the univariate approach ( $2.672 \pm 0.060$ ). The linear multitrait model cannot be applied to nearly the same number of traits as the approximate model. Therefore, the approximate model with developments handling breeding values from more complex models than presented in this paper is an option of choice in countries providing total merit indices that combine many traits because it does not neglect correlations between these traits.

**Key words:** stochastic simulation, multitrait model, total merit

### INTRODUCTION

In the dairy industry, most countries have total merit indices combining information from different traits to achieve greater economic returns through the breeding program (Miglior et al., 2005). This is, in general, done by combining EBV for the individual traits without taking into account the correlation between traits and differences in accuracy of EBV for the traits. Within production traits, correlations between milk, fat, and protein yield are often taken into account when predicting breeding values. With the introduction of functional traits such as fertility, longevity, resistance to mastitis, and other diseases, this approach is questionable (Lindhé and Phillipson, 1998; Ducrocq et al., 2001). These traits have low heritabilities, resulting in low accuracy on EBV, especially for cows and young bulls undertesting when they get their first proofs. Therefore, decisions on early selection for these traits are associated with uncertainty.

Another problem in doing multitrait selection based on breeding values from single trait analyses is that breeding values are probably biased because early selection decisions are not accounted for (Ducrocq et al., 2003). An important condition for BLUP EBV to be unbiased is that all selection decisions are included in the data, and this will not be the case in single trait evaluations when the selection is on a total merit index (Henderson, 1975; Schaeffer et al., 1998; Ducrocq et al., 2003; Mrode, 2005).

Even though breeding values for functional traits have been provided in the Danish dairy cattle breeding system for more than 2 decades, known selection for these traits has not been emphasized as much as the economic value of these traits implies. Instead, selection focus has been placed mainly on production and type traits. This has resulted in a negative genetic trend for many functional traits in the breeding goal (Danish Cattle, 2006).

A large number of traits are recorded for breeding purposes in the dairy industry and a number of these are both phenotypically and genetically correlated to

Received July 6, 2006.

Accepted February 8, 2007.

<sup>1</sup>Corresponding author: jan.lassen@agrsci.dk

many of the cost-related traits (e.g., Short and Lawlor, 1992; Rupp and Boichard, 1999; Sørensen et al., 2000; Haile-Mariam et al., 2004). Using information from these indicator traits, such as conformation traits or SCS, can significantly increase accuracy of total merit indices (Ducrocq et al., 2001; Lassen et al., 2003).

Multiple-trait BLUP evaluations combining direct and indirect information have been suggested to circumvent these problems. Their properties, benefits, and limitations compared with univariate models are well known (Thompson and Meyer, 1986; Van der Werf et al., 1992; Villanueva et al., 1993; Ducrocq, 1994). Although theoretically possible, implementations of these models for practical use are still unrealistic. Even though computer power has increased greatly in recent years, the amount of information to process is still a bottleneck. Therefore, approximations are needed. These approximations have been made with different approaches using deregressed proofs, daughter yield deviations, or preadjusted data (Schaeffer, 1994; Colleau et al., 1999; Druet et al., 1999; Ducrocq et al., 2001). These approaches have been implemented in both national and international evaluation procedures, and they make it possible to account for the fact that the traits under selection are indeed correlated. The multiple-trait across-country genetic evaluation (MACE) method of Schaeffer (1994) does not allow residual correlations to differ from zero, which is not a problem when records of animals from different countries are used. When correlated traits are measured on the same animal, this is no longer valid. However, recent developments in MACE now allow it to account for residual correlations between traits (Sullivan and Wilton, 2001; Sullivan et al., 2005). Another way to do multitrait evaluation is via the use of preadjusted data in a 2-step procedure. The first step is an adjustment of the phenotypes by subtracting the estimated fixed effects and nongenetic random effects obtained from single trait evaluation (Ducrocq et al., 2001). The second step is a multitrait evaluation on the preadjusted records, including residual correlations.

In this study the objective was to validate an approximate multitrait evaluation using preadjusted data with linear models so that systems were manageable and comparable. The focus was to compare the outcome of using an approximate multitrait model that was computationally feasible to a full linear model that was not computationally feasible on large data sets for national breeding value estimation. This was done in a stochastic simulation study of a dairy cattle population in which gains in total merit combining a limited number of traits were compared over 4 different approaches.

**Table 1.** Parameters for the population used in the simulation study

| Parameter  |        |
|--|--------|
| Number of years simulated                                | 15     |
| Number of replicates                                     | 20     |
| Number of herds  | 1,000  |
| Herd size (number of females)                            | 200    |
| Young sires used for mating (%)                          | 30     |
| Semen doses used from each young sire                    | 750    |
| Semen doses from each proven sire per year               | 24,000 |
| Maximum age of bull at mating (yr)                       | 7      |
| Age when observations from females become available (mo) | 33     |
| Maximum number of offspring per dam                      | 6      |
| Pregnancy rate (%)                                       | 0.9    |

## MATERIALS AND METHODS

### Outline of the Simulation

The simulation was in 2 steps each covering 15 yr of selection. The idea behind this setup was to create a population in the first step that had been undergoing selection for a number of years before applying different models for prediction of breeding values and different breeding goals in the second step. Between the 2 steps, (co)variance components for the traits were estimated and used for prediction of EBV in the last step. This corresponded to a situation that often occurs in practical breeding programs in which (co)variance components are reestimated when new methods or models for prediction of EBV are introduced and the breeding goals are changed.

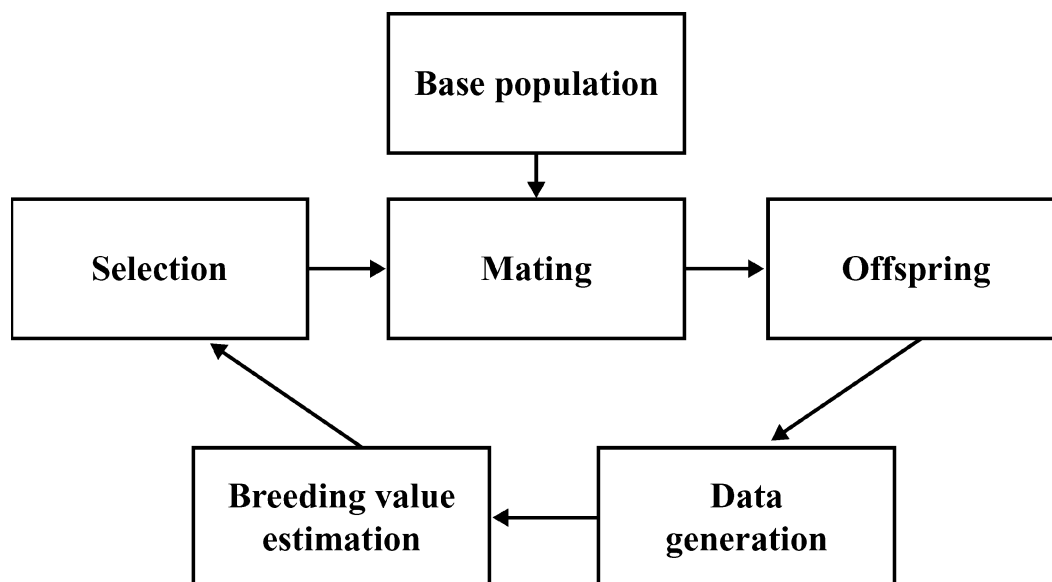
In the second step, 20 replicates of 4 scenarios were simulated, all starting from the same population. Comparison of the different scenarios was based on the genetic trend for total merit and the individual traits that contributed to the total merit index were compared.

### Population Structure

A population of approximately 100,000 animals was simulated using a modified version of the stochastic simulation program DairySim (Sørensen et al., 1999). The simulation was performed so that the population generated mimicked a dairy cattle population (Table 1) and followed a yearly cycle as shown in Figure 1.

### Traits

Seven traits were chosen to represent different traits of interest in a total merit breeding goal (Table 2). There were 5 normally distributed (linear) traits and 2 binary traits: milk production (**Y**, linear), udder depth (**UD**, linear), mastitis occurrence (**MO**, binary), nonreturn rate (**NR**, binary), dairy form (**DF**, linear), days open (**DO**, linear), and SCS (linear). At generation zero, the genetic and phenotypic parameters for these traits were



**Figure 1.** One year in the simulation study.

known and considered to be the true ones (Table 3). For all traits the phenotypic variance was set to 1, meaning that the genetic variance was equal to the heritability.

These traits were chosen both to represent traits with considerable economic value in practical breeding and because they represent different types of traits with a wide range of genetic variance and correlation with other traits. The parameters were obtained from an intensive literature study, and the genetic (co)variance matrix was positive definite. Breeding values and phenotypic observations are simulated multivariate as Gaussian traits, but MO and NR are converted to binary traits with thresholds corresponding to frequencies of 0.18 and 0.5.

### **The First Fifteen Years**

Breeding values and phenotypes for the 7 traits were simulated for base population animals, and they were allocated to age classes, gender, and herd-year-season. The observations were simulated using the information from the parameters in Table 3. In the base population, the animals were mated randomly. In later years, the

best males and females based on predicted total merit index (**TMI**) were mated to produce offspring. The TMI use was based on EBV for Y and MO with economic weights of 19.4 and –50.0, respectively, corresponding to the economic values currently used in the Danish Holstein population (Danish Agricultural Advisory Centre, 2006). The EBV for Y were obtained from a single trait animal model, and the EBV for MO were from a trivariate animal model for MO, UD, and SCS. Selected bulls could be used in all herds, but to avoid strong inbreeding, mating between full- and half-sibs was not allowed. Bulls were allocated to 1 of 2 groups representing young bulls and proven bulls. This was done to mimic the current use of bulls as closely as possible. Offspring were simulated individually and randomly distributed on gender and for each individual trait allocated to a herd-year-season. True breeding values were simulated as their parents' average breeding values plus Mendelian sampling terms. Residuals and observations were also simulated. When females reached the age of 35 mo, their phenotypes were realized. Observations from animals that did not reach this age were discarded. Cows were kept in the herds until they were involuntary culled, until better replacement

**Table 2.** Name of traits, their type, and abbreviation in the simulation study

| Trait        | Milk production | Udder depth | Mastitis occurrence | Nonreturn rate | Dairy form | Days open | SCS    |
|--------------|-----------------|-------------|---------------------|----------------|------------|-----------|--------|
| Type         | Linear          | Linear      | Binary              | Binary         | Linear     | Linear    | Linear |
| Abbreviation | Y               | UD          | MO                  | NR             | DF         | DO        | SCS    |

**Table 3.** True heritabilities (diagonal, in bold), genetic (below diagonal), and phenotypic (above diagonal) correlations for traits in the simulation study

| Trait <sup>1</sup> | Y           | UD          | MO          | NR          | DF          | DO          | SCS         |
|--------------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| Y                  | <b>0.30</b> | -0.20       | 0           | -0.10       | 0.25        | 0.20        | -0.15       |
| UD                 | -0.35       | <b>0.30</b> | -0.05       | 0           | 0           | -0.10       | -0.05       |
| MO                 | 0.35        | -0.60       | <b>0.04</b> | 0           | 0           | 0           | 0.20        |
| NR                 | -0.35       | 0.30        | -0.10       | <b>0.03</b> | 0           | 0.05        | 0           |
| DF                 | 0.45        | -0.10       | 0.25        | 0           | <b>0.25</b> | 0.10        | 0           |
| DO                 | 0.55        | -0.10       | 0.05        | -0.10       | 0.45        | <b>0.04</b> | 0           |
| SCS                | 0.15        | -0.30       | 0.75        | -0.20       | 0.25        | -0.25       | <b>0.10</b> |

<sup>1</sup>Y = milk production; UD = udder depth; MO = mastitis occurrence; NR = nonreturn rate; DF = dairy form; DO = days open.

heifers were available, or until they reached the maximum age of 8 yr.

### Estimation of (Co)variance Components

After the first 15 yr, (co)variance components to be used in prediction of EBV for the last 15 yr were estimated. Four scenarios were conducted: **N1**, **B1**, **B2**, and **B3**, where N and B indicate a narrow or a broad breeding goal, respectively, and 1, 2, and 3 indicate which model was used for prediction of breeding values. Estimation of the (co)variance components for each of the 4 scenarios were performed on 10 sample datasets each with ~8,000 animals with data. A sampling procedure and an animal model were chosen to take better account of selection among female individuals in the population. For scenarios N1, B1, and the first step of B2, (co)variance components were estimated using 3 single-trait animal models for Y, NR, and DO, and a 3-trivariate animal model for MO and UD with SCS. For the last step of scenario B2, (co)variance components were estimated using a full multivariate animal model on preadjusted data and keeping the variances fixed at the values estimated for the first step. For scenario B3, a full multitrait animal model was used. The average over the 10 subsets was used as parameters for prediction of EBV for the last 15 yr. The estimations were conducted with the AI-REML module of the DMU package (Madsen and Jensen, 2005).

### The Last Fifteen Years

After step 1 and estimation of (co)variance components, 20 replicates of each of the 4 scenarios were simulated. All replicates of all scenarios were started from the same simulated population.

Scenario N1 was a continuation of the setup for the first 15 yr but with reestimated (co)variance components. All of the 3 other scenarios had a broader breeding goal with weights on Y (19.4), MO (-50.0), UD (4.2), NR (13.0), and DO (-16.75). The 3 scenarios differ in

the model for prediction of breeding values. In scenario B1, the EBV were estimated in 3 single trait animal models for Y, NR, and DO, and a trivariate animal model for MO and UD with SCS as an additional trait of information. Scenario B2 used an approximate multitrait model as proposed by Ducrocq et al. (2001) and Besbes et al. (2002). This is a 2-step procedure, where the first step applied models corresponding to the model used in scenario N1 and B1. Based on estimates from these models, data were adjusted for all effects other than the genetic effect. Using these preadjusted data, breeding values for all traits were predicted in a multitrait setting using a model with a mean, a genetic effect, and a residual. The last scenario (B3) used a full multitrait animal model on raw data for prediction of EBV.

For the final 15 yr, the population was simulated following the same concept as described for the first 15 yr, although 4 different approaches were used to perform the breeding value estimation, with 4 different sets of estimated genetic ( $G_0$ ) and residual ( $R_0$ ) (co)variance matrices. In approach N1, the following model corresponding to the model used during the first 15 yr was used:

$$y_{i,m} = hys_i + a_{i,m} + e_{i,m} \quad [1]$$

where  $y_{i,m}$  represents the record for animal m for trait i,  $hys_i$  is the herd-year-season effect for each trait, and  $a_{i,m}$  and  $e_{i,m}$  are the additive genetic and residual term for trait i of animal m. This is done using the following (co)variance structure to predict breeding values:

$$\text{Var} \begin{pmatrix} a \\ e \end{pmatrix} \sim N \left( \mathbf{0}; \begin{bmatrix} \mathbf{G}_0 \otimes \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_0 \otimes \mathbf{I} \end{bmatrix} \right). \quad [2]$$

In N1, 4 traits were analyzed, so  $G_0$  and  $R_0$  were structures with a  $1 \times 1$  matrix for the univariate analysis of Y and a  $3 \times 3$  matrix for the analysis of MO, UD, and SCS (Equation 2). This matrix was estimated using



equation 1. In B1, 7 traits were analyzed, so  $G_0$  and  $R_0$  were structures with a  $1 \times 1$  matrix for Y, NR, DF, and DO, and a  $3 \times 3$  matrix for the analysis of MO, UD, and SCS. The matrix A defined the genetic relationship among animals and the matrix I was an identity matrix of proper order. Equivalently, the following (co)variance structure was used to obtain EBV:

$$\begin{bmatrix} h_p^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & h_{ud}^2 & \sigma_{e(ud,mo)} & 0 & 0 & 0 & \sigma_{e(ud,scs)} \\ 0 & \sigma_{g(ud,mo)} & h_{mo}^2 & 0 & 0 & 0 & \sigma_{e(mo,scs)} \\ 0 & 0 & 0 & h_n^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & h_{df}^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & h_{do}^2 & 0 \\ 0 & \sigma_{g(ud,scs)} & \sigma_{g(mo,scs)} & 0 & 0 & 0 & h_{scs}^2 \end{bmatrix}$$

where heritabilities are on the diagonal, residual (co)variances above, and additive genetic (co)variances below the diagonal. In approach B2 the data was adjusted for the fixed effects in a 2-step procedure (equation 3). The fixed effects were estimated from a model corresponding to the model used in approach B1 so each trait only contained a mean, a genetic, and a residual effect (Ducrocq et al., 2001; Besbes et al., 2002). Correcting the data for fixed effects led to preadjusted data, which were used to predict breeding values for all traits in a multitrait setting; as follows:

$$y_{i,m} = hys_i + a_{i,m} + e_{i,m} \rightarrow y_{i,m}^* = \mu_i + a_{i,m}^* + e_{i,m}^* \quad [3]$$

where  $y_{i,m}$  represents the record for animal m for trait i,  $hys_i$  is the herd-year-season effect for each trait,  $a_{i,m}$  and  $e_{i,m}$  are the additive genetic and residual term for trait i of animal m,  $y_{i,m}^*$  is the preadjusted record for animal m for trait i,  $\mu_i$  is the overall mean for each trait, and  $a_{i,m}^*$  and  $e_{i,m}^*$  are the additive genetic and residual terms for the preadjusted record i of animal m, respectively. The second step in equation 2 was a complete multitrait setting, where  $G_0$  and  $R_0$  were  $7 \times 7$  matrices. This configuration yielded the following (co)variance structure to predict breeding values:

$$\begin{bmatrix} h_p^2 & \sigma_{e(p,ud)}^* & \sigma_{e(p,mo)}^* & \sigma_{e(p,nr)}^* & \sigma_{e(p,df)}^* & \sigma_{e(p,do)}^* & \sigma_{e(p,scs)}^* \\ \sigma_{g(p,ud)}^* & h_{ud}^2 & \sigma_{e(ud,mo)}^* & \sigma_{e(ud,nr)}^* & \sigma_{e(ud,df)}^* & \sigma_{e(ud,do)}^* & \sigma_{e(ud,scs)}^* \\ \sigma_{g(p,mo)}^* & \sigma_{g(ud,mo)}^* & h_{mo}^2 & \sigma_{e(mo,nr)}^* & \sigma_{e(mo,df)}^* & \sigma_{e(mo,do)}^* & \sigma_{e(mo,scs)}^* \\ \sigma_{g(p,nr)}^* & \sigma_{g(ud,nr)}^* & \sigma_{g(mo,nr)}^* & h_{nr}^2 & \sigma_{e(nr,df)}^* & \sigma_{e(nr,do)}^* & \sigma_{e(nr,scs)}^* \\ \sigma_{g(p,df)}^* & \sigma_{g(ud,df)}^* & \sigma_{g(mo,df)}^* & \sigma_{g(nr,df)}^* & h_{df}^2 & \sigma_{e(df,do)}^* & \sigma_{e(df,scs)}^* \\ \sigma_{g(p,do)}^* & \sigma_{g(ud,do)}^* & \sigma_{g(mo,do)}^* & \sigma_{g(nr,do)}^* & \sigma_{g(df,do)}^* & h_{do}^2 & \sigma_{e(do,scs)}^* \\ \sigma_{g(p,scs)}^* & \sigma_{g(ud,scs)}^* & \sigma_{g(mo,scs)}^* & \sigma_{g(nr,scs)}^* & \sigma_{g(df,scs)}^* & \sigma_{g(do,scs)}^* & h_{scs}^2 \end{bmatrix}$$

where heritabilities are on the diagonal, residual (co)variances above, and additive genetic (co)variances below the diagonal. Note that variances were kept fixed at the estimates obtained from N1. Finally, approach B3 was a complete multitrait model (equation 1), where fixed and random effects were estimated at the same time and where  $G_0$  and  $R_0$  were  $7 \times 7$  matrices (equation 2). That approach used the same (co)variance structure used to predict breeding values from the data as already described for the preadjusted data, but variances were not fixed.

## RESULTS

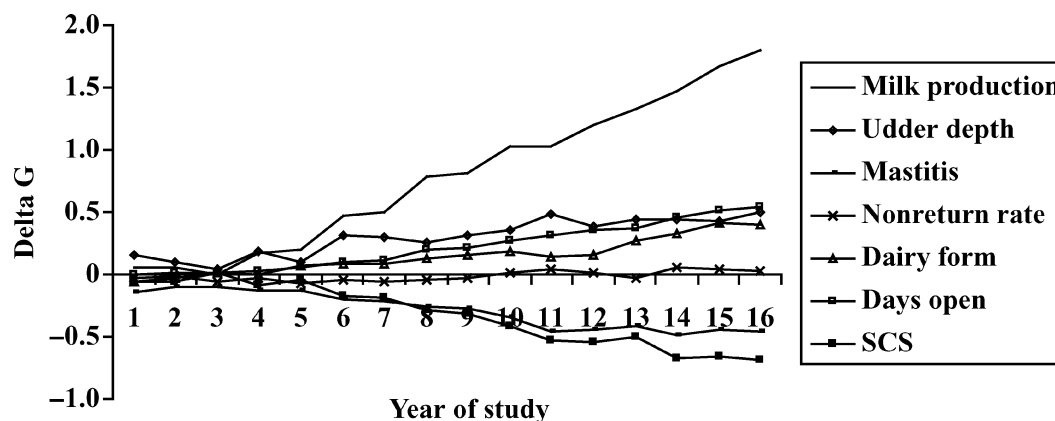
### Genetic Trend for the First Fifteen Years

The genetic trends for the first period in the replicate used as the base for the second period are shown in Figure 2. A substantial increase in genetic gain was obtained for Y, and at the same time, it was possible to obtain an improvement (i.e., a decrease) for MO and SCS. Nonreturn rate stayed more or less at the same level during the 15-yr period, whereas DO were markedly increased. Udder depth and DF both had a positive trend in the period. Using the results shown gives a regression of total merit on year of birth of 1.81 using the economic weights from breeding goal N.

### Genetic Trend for the Final Fifteen Years

Scenario B3 gave the highest true genetic progress in total merit (Table 4). Scenario B3 was superior to scenario B2, and these 2 approaches were both significantly better than scenario B1. With the same amount of data and information, scenario B2 increased the total merit by 5.5% compared with scenario B1, and comparing scenario B3 with scenario B1, the increase was 15.0% for progress in total merit. Scenario N1 gave the lowest progress in total merit. The value for scenario N1 (2.35) is somewhat lower than the regression coefficient obtained in the first 15 yr (2.47). This was expected because estimated genetic parameters were used instead of the true ones. Comparing scenarios N1 and B1, a significant effect of changing the breeding goal using the same model for prediction of breeding values was observed. The predicted genetic trends for total merit were all less than their observed values, and this difference decreased with more advanced models.

All scenarios gave a large increase in true genetic merit for Y (Table 5). Scenario N1 with the highest relative weight on Y also gave the highest increase. Given that the model used in scenario B3 was the optimal model, scenario N1 gave a 30.2% higher increase in Y than scenario B3. Scenarios B1, B2, and B3 all had similar progress for Y and they were all signifi-



**Figure 2.** Average genetic merit (Delta G) in the 7 individual traits for cows born in the respective years of the study.

cantly lower than scenario N1. There was a decrease in UD score only in scenario N1. Scenario B3 had a significantly higher genetic merit for MO than scenarios N1 and B1, and scenario N1 was significantly lower than the other 3 scenarios. For MO, only 8.6 and 65.2% of the genetic progress obtained in scenario B3 was obtained in scenarios N1 and B1, respectively. For the 2 reproduction traits (NR and DO), a genetic decrease was observed in the last 15 yr in all scenarios, except for NR in scenario B3. However, for NR the trend was not significantly different from zero except for scenario N1. For reproduction traits, scenario N1 was significantly worse than the other scenarios, and the decrease in genetic merit was smaller with more advanced models. Compared with the genetic progress in scenario B3, the relative efficiency for DO in scenario N1 was 51.5%, whereas it was 89.5 and 100% for scenarios B1 and B2, respectively. For SCS, all models showed an improvement of the genetic merit, but there was no obvious pattern between scenarios. In general, there was very little difference between individual traits, but these differences summed up to a significant difference in total merit.

The trends for EBV from the 7 individual traits were generally very close to the true breeding values (Table

6). For the production scenario, N1 was best—16.1% greater than scenario B3. For MO, the difference between true and predicted breeding values was relatively larger, and this difference was the primary cause of the somewhat lower predicted breeding values for total merit compared with the true total merit breeding values. In terms of relative efficiency, only 28.5, 50.0, and 64% of the genetic improvement obtained in scenario B3 were obtained in scenarios N1, B1, and B2, respectively. A reason for this could be the relatively large change in the estimated genetic correlations between MO and the other traits in the breeding goal compared with the ones used to simulate the data. Another reason could be the use of linear models for the analysis of a discrete trait with a low frequency. As for the true breeding values, it was not possible to observe an improvement for reproduction, and none of the scenarios were significantly different from zero for NR. Compared with MO, which was also simulated as a binary trait, the estimated genetic correlations with the other traits in the breeding goal did not change as much for NR (Tables 7, 8, and 9). Also, the frequency for N was closer to 0.5, which fit better with a linear assumption of the trait. That might be the main reason why N did not change as much between models as did MO. In scenario

**Table 4.** Average regression coefficients of true and predicted genetic values of cows for total merit on year of birth for the 4 different scenarios using the economic values from breeding goal B with corresponding empirical SE over 20 replicates

| Value     | Scenario <sup>1</sup> |               |               |               |
|-----------|-----------------------|---------------|---------------|---------------|
|           | N1                    | B1            | B2            | B3            |
| True      | 1.706 (0.060)         | 2.672 (0.060) | 2.819 (0.047) | 3.073 (0.069) |
| Predicted | — <sup>2</sup>        | 2.325 (0.084) | 2.566 (0.077) | 3.021 (0.046) |

<sup>1</sup>N = narrow breeding goal; B = broad breeding goal; 1 = using a single-trait model for milk production, nonreturn rate, dairy form, and days open, and a multitrait model for udder depth, mastitis occurrence, and SCS; 2 = using an approximate multitrait model; and 3 = using a full multitrait model.

<sup>2</sup>In scenario N1, breeding values for nonreturn rate, dairy form, and days open were not predicted.

**Table 5.** Average regression coefficients of true genetic values of cows for the 7 individual traits on year of birth for the first 15 yr and for each of the 4 different scenarios<sup>1</sup> with corresponding SE

| Trait               | Desired direction | First 15 yr | Scenario <sup>1</sup> |                |                |                |
|---------------------|-------------------|-------------|-----------------------|----------------|----------------|----------------|
|                     |                   |             | N1                    | B1             | B2             | B3             |
| Milk production     | +                 | 0.111       | 0.138 (0.005)         | 0.109 (0.008)  | 0.097 (0.007)  | 0.106 (0.007)  |
| Udder depth         | +                 | 0.025       | -0.014 (0.009)        | 0.029 (0.005)  | 0.043 (0.012)  | 0.030 (0.008)  |
| Mastitis occurrence | -                 | -0.015      | -0.002 (0.005)        | -0.015 (0.006) | -0.021 (0.006) | -0.023 (0.004) |
| Nonreturn rate      | +                 | 0.001       | -0.019 (0.002)        | -0.003 (0.004) | -0.002 (0.004) | 0.001 (0.003)  |
| Dairy form          | 0                 | 0.025       | 0.050 (0.005)         | 0.039 (0.008)  | 0.035 (0.010)  | 0.032 (0.007)  |
| Days open           | -                 | 0.039       | 0.033 (0.004)         | 0.019 (0.004)  | 0.017 (0.003)  | 0.017 (0.004)  |
| SCS                 | -                 | -0.026      | -0.016 (0.005)        | -0.022 (0.005) | -0.022 (0.004) | -0.018 (0.005) |

<sup>1</sup>N = narrow breeding goal; B = broad breeding goal; 1 = using a single-trait model for milk production, nonreturn rate, dairy form, and days open, and a multitrait model for udder depth, mastitis occurrence, and SCS; 2 = using an approximate multitrait model; and 3 = using a full multitrait model.

N1, there was a slight setback for genetic merit for UD. In scenario N1, breeding values for NR, DF, and DO were not predicted.

The estimation of (co)variances after 15 yr showed rather large variability among the different models applied, especially when looking at the correlations between Y and the other individual traits (Tables 7, 8, and 9). The genetic correlations obtained from the full multitrait model showed the least discrepancy from the genetic correlations used to simulate the data. The genetic correlations obtained in the approximate multitrait model were generally lower than the ones used to simulate the data.

The breeding value estimation in the last round of the simulation took 15% more CPU time using the full multitrait model in scenario B3 than for the approximate multitrait model used in scenario B2. In each of the 4 scenarios, more than 2 gigabytes of data was generated from one replicate given the setup of the simulations. With 20 replicates the result was more than 160 gigabytes of data.

## DISCUSSION

The purpose of doing 15 yr of selection before applying the new models was not to quantify the genetic development in that period, but to create a population with a genetic structure that had undergone selection for a period in correspondence to a real dairy cattle population where the models will be applied. Because of a computing time of approximately 80 h per replicate, only 20 replicates were considered.

The full multitrait model gave the greatest genetic progress in total merit. The approximate multitrait model performed well, however, and provided significantly greater genetic progress for total merit than with univariate models. The genetic progress for the individual traits was changed: the contribution to the total genetic merit came more from cost-reducing traits and less from production when a more complex model was used. This agrees with the results of Ducrocq et al. (2001) and Besbes et al. (2002), who applied similar methods on real data; they obtained a higher and more sustainable breeding profile in their total merit index,

**Table 6.** Average regression coefficients of predicted genetic values of cows for the 7 individual traits on year of birth for the 4 different scenarios<sup>1</sup> with corresponding SE

| Trait               | Desired direction | Scenario <sup>1</sup> |                |                |                |
|---------------------|-------------------|-----------------------|----------------|----------------|----------------|
|                     |                   | N1                    | B1             | B2             | B3             |
| Milk production     | +                 | 0.137 (0.008)         | 0.106 (0.008)  | 0.113 (0.009)  | 0.118 (0.006)  |
| Udder depth         | +                 | -0.007 (0.007)        | 0.027 (0.006)  | 0.027 (0.011)  | 0.040 (0.008)  |
| Mastitis occurrence | -                 | -0.004 (0.001)        | -0.007 (0.002) | -0.009 (0.003) | -0.014 (0.002) |
| Nonreturn rate      | +                 | — <sup>2</sup>        | -0.001 (0.001) | -0.001 (0.002) | 0.003 (0.001)  |
| Dairy form          | 0                 | —                     | 0.036 (0.007)  | 0.032 (0.009)  | 0.032 (0.006)  |
| Days open           | -                 | —                     | 0.012 (0.002)  | 0.016 (0.003)  | 0.011 (0.003)  |
| SCS                 | -                 | -0.025 (0.009)        | -0.020 (0.004) | -0.023 (0.005) | -0.030 (0.006) |

<sup>1</sup>N = narrow breeding goal; B = broad breeding goal; 1 = using a single-trait model for milk production, nonreturn rate, dairy form, and days open, and a multitrait model for udder depth, mastitis occurrence, and SCS; 2 = using an approximate multitrait model; and 3 = using a full multitrait model.

<sup>2</sup>In scenario 1, breeding values for nonreturn rate, dairy form, and days open were not predicted.



**Table 7.** Heritabilities (on diagonal, in bold), genetic (below diagonal), and residual (above diagonal) correlations between the 7 traits in the simulation estimated for use in scenarios N1 and B1<sup>1</sup>

| Trait <sup>2</sup> | Y           | UD              | MO          | NR          | DF          | DO          | SCS         |
|--------------------|-------------|-----------------|-------------|-------------|-------------|-------------|-------------|
| P                  | <b>0.27</b> | NE <sup>3</sup> | NE          | NE          | NE          | NE          | NE          |
| UD                 | NE          | <b>0.29</b>     | 0.03        | NE          | NE          | NE          | 0.10        |
| MO                 | NE          | -0.41           | <b>0.05</b> | NE          | NE          | NE          | 0.09        |
| NR                 | NE          | NE              | NE          | <b>0.03</b> | NE          | NE          | NE          |
| DC                 | NE          | NE              | NE          | NE          | <b>0.25</b> | NE          | NE          |
| DO                 | NE          | NE              | NE          | NE          | NE          | <b>0.06</b> | NE          |
| SCS                | NE          | -0.23           | 0.51        | NE          | NE          | NE          | <b>0.12</b> |

<sup>1</sup>N = narrow breeding goal; B = broad breeding goal; 1 = using a single-trait model for milk production, nonreturn rate, dairy form, and days open, and a multitrait model for udder depth, mastitis occurrence, and SCS. The SE on the heritabilities were below 0.01, and SE on correlations were all below 0.05.

<sup>2</sup>Y = milk production; UD = udder depth; MO = mastitis occurrence; NR = nonreturn rate; DF = dairy form; DO = days open.

<sup>3</sup>NE = correlation was not estimated.

with less increase in production traits and more progress, or at least less decline, in a number of cost reducing traits. The genetic progress obtained by Ducrocq et al. (2001) was partly due to a change in the breeding goal at the same time as implementing the approximate multitrait model. In this study, estimating new genetic parameters after 15 yr of selection seemed to be the main element influencing the results, because the multitrait model corrects for selection bias due to ignoring the genetic correlation between production and the other traits.

Within breeding goal, no significant differences were observed between scenarios for the individual traits. Using a more advanced model led to a more favorable genetic trend for the cost-reducing traits, and this led to significant difference in total merit. However, a more advanced model for prediction of breeding values did not increase progress for production.

In recent years, much effort has been expended into developing test-day models for analysis of production traits. Methodology to implement breeding values from a test-day model in an approximate multitrait model, as described in this paper, and in a MACE setting has

been developed (Liu et al., 2004; Tarrés et al., 2006b). Such approaches were not used in this study, however. To use this as the method of this study, a simple estimate of genetic merit for production is needed even if a test-day model is used. One possibility is to calculate yield deviations as the sum of the vector of random regression coefficients times the orthogonal elements of the model and equivalent daughter yield deviations (Mrode, 2005). This can be done to create pseudo-305-d records for production to use in the approximate multitrait model.

Although the multitrait model performed better than the approximate model, such a model is not computationally feasible for data sets that are much larger than the data set simulated in this study. The purpose of comparing the approximate model to the full linear multitrait model was to assess the efficiency of the approximate model to the optimum. The total merit index in Denmark consists of a large number of traits. Therefore, clustering traits into groups of traits might be a useful implementation of the approximate 2-step multitrait model, as may ignoring correlations that are numerically small, such as <0.10 in absolute value.

**Table 8.** Heritabilities (on diagonal, in bold), genetic (below diagonal), and residual (above diagonal) correlations between the 7 traits in the simulation estimated for use in scenario B2<sup>1</sup>

| Trait <sup>2</sup> | Y           | UD          | MO          | NR          | DF          | DO          | SCS         |
|--------------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| Y                  | <b>0.27</b> | -0.17       | -0.05       | -0.08       | 0.17        | 0.21        | -0.21       |
| UD                 | -0.38       | <b>0.29</b> | 0.03        | -0.05       | -0.24       | -0.14       | 0.02        |
| MO                 | 0.42        | -0.39       | <b>0.05</b> | -0.01       | -0.04       | -0.00       | 0.09        |
| NR                 | -0.27       | 0.20        | -0.15       | <b>0.03</b> | 0.00        | 0.04        | 0.02        |
| DF                 | 0.48        | -0.16       | 0.24        | 0.12        | <b>0.25</b> | 0.06        | 0.02        |
| DO                 | 0.32        | -0.17       | 0.11        | -0.16       | 0.34        | <b>0.06</b> | 0.020       |
| SCS                | 0.16        | -0.24       | 0.68        | -0.19       | -0.07       | -0.33       | <b>0.12</b> |

<sup>1</sup>B2 = broad breeding goal using an approximate multitrait model. Heritabilities were estimated with univariate models and kept fixed in the multitrait setting. The SE on the heritabilities were below 0.01, and SE on correlations were all below 0.06.

<sup>2</sup>Y = milk production; UD = udder depth; MO = mastitis occurrence; NR = nonreturn rate; DF = dairy form; DO = days open.

**Table 9.** Heritabilities (on diagonal, in bold), genetic (below diagonal), and residual (above diagonal) correlations between the 7 traits in the simulation estimated for use in scenario B3<sup>1</sup>

| Trait <sup>2</sup> | Y           | UD          | MO          | NR          | DF          | DO          | SCS         |
|--------------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| Y                  | <b>0.27</b> | -0.13       | -0.06       | -0.08       | 0.15        | 0.19        | -0.20       |
| UD                 | -0.41       | <b>0.33</b> | 0.03        | -0.05       | -0.20       | -0.12       | 0.01        |
| MO                 | 0.39        | -0.38       | <b>0.05</b> | -0.01       | -0.05       | -0.01       | 0.11        |
| NR                 | -0.36       | 0.16        | -0.12       | <b>0.04</b> | 0.00        | 0.04        | 0.02        |
| DF                 | 0.47        | -0.21       | 0.24        | 0.08        | <b>0.24</b> | 0.05        | 0.01        |
| DO                 | 0.34        | -0.19       | 0.16        | -0.17       | 0.32        | <b>0.05</b> | 0.02        |
| SCS                | 0.18        | -0.23       | 0.75        | -0.16       | -0.02       | -0.31       | <b>0.11</b> |

<sup>1</sup>B3 = broad breeding goal using a full multitrait model. The SE on the heritabilities were below 0.02, and SE on correlations were all below 0.05.

<sup>2</sup>Y = milk production; UD = udder depth; MO = mastitis occurrence; NR = nonreturn rate; DF = dairy form; DO = days open.

Such an approach would probably still be better than ignoring correlations between traits completely, which would lead to low accuracies of EBV, selection bias, and less genetic progress.

For the individual traits, scenario N1 is clearly best for Y. This is mainly due to the relative higher economic weight put on the trait in this scenario. In general, the genetic trends depend a lot on the economic weights that have been applied. The weights used are similar to those used in the Danish breeding value system for Holsteins.

Substantial differences in the genetic trends for udder depth were observed between scenario N1 and the other scenarios (Table 5). In particular, for scenario B1, with a nonzero economic weight on UD, a positive trend for the trait was obtained.

In this study, increases for NR or DO were observed only in scenario B3, and the genetic change was not significantly different from zero. In fact, few studies have been able to show important improvements in genetic gain for reproduction, but surely the results obtained in this study are also a consequence of the economic weights used. Days open is very highly correlated to Y and DF, and the positive weight put on Y in this study made it practically impossible to improve DO.

No differences were observed for SCS among the scenarios. The trait served only as an indicator trait in this study and its genetic trend was mainly influenced by the correlations to MO and Y and the weights on these 2 traits. The selection for less MO improved SCS, whereas the selection for increased Y made it decline.

Dairy cattle breeders face the challenge that many traits with significant economic value have an unfavorable genetic correlation with milk yield, and yield remains the most important trait in the breeding goal. Using multitrait models and taking into account these unfavorable genetic correlations makes it possible to improve production simultaneously with other traits of economic importance. Multitrait selection using single trait EBV is currently practiced in many countries. This

study shows that this procedure not only leads to less genetic progress, but also results in EBV that are biased due to the fact that selection on correlated traits is ignored.

Although this study was based on linear models, the same approach can also be applied with nonlinear, binary, and survival traits (Ducrocq et al., 2001; Besbes et al., 2002; Tarrés et al., 2006a). The focus of this study was not to examine the effect of implementing different models on different traits, but to quantify the efficiency of the model proposed by Ducrocq et al. (2001) against a full linear model. Use of linear models for evaluation of nonlinear traits is still common worldwide. Linear models might also be the first choice when a full multitrait analysis is computationally feasible to apply. Therefore, this study focused on comparison of the approximate multitrait model to a full linear model.

The use of preadjusted data in a 2-step procedure has never been formally compared with the use of daughter yield deviation (DYD) in a MACE procedure where residual correlations are taken into account. In some ways the 2 procedures are similar, however. For traits with one observation per animal, the deregressed proof or DYD of a trait is the same as correcting a phenotype for fixed effects. Methodology has been developed to include EBV from test-day models, threshold models, and survival models for both methods. In the MACE procedure, effective daughter contributions are used to take account for the different amount of information from each animal, whereas the 2-step procedure uses the accuracy obtained from the univariate evaluations as weights in the multitrait evaluation. Lately, a lot of effort has been put into including EBV from a test-day model into a MACE setting using DYD and a multiple effective daughter contribution methodology (Liu et al., 2004; Tarrés et al., 2006b). In this study, where only one observation per trait was simulated, using a MACE procedure as described by Sullivan and Wilton (2001) would lead to the same results as using the 2-step procedure as described by Ducrocq et al. (2001).

## CONCLUSIONS

When evaluating and selecting for many correlated traits, using an approximate multitrait model for prediction of breeding values will yield greater genetic progress for total merit than will using EBV from individual univariate models. The difference in response in the individual traits appears negligible, but in total the effect on total merit can be substantial. The approximate model yields less selection response than does using EBV from a full linear multitrait model, but has the advantage of being computationally feasible with the large data sets typical of national selection programs with large numbers of traits.

## REFERENCES

- Besbes, B., V. Ducrocq, and M. Protais. 2002. An approximate total merit index combining linear traits, a survival trait and a categorical trait in laying hens. *Proc. 7th World Congr. Genet. Appl. Livest. Prod.* CD-ROM communication No. 20-05.
- Colleau, J. J., V. Ducrocq, D. Boichard, and H. Larroque. 1999. Approximate multi-trait BLUP evaluation to combine functional traits information. Final Genetic Improvement of Functional Traits in Cattle Workshop. Breeding Goals and Selection Schemes. Ede-Wageningen, The Netherlands. *Interbull Bull.* 23:151-160.
- Danish Agricultural Advisory Centre. 2006. <http://www.lr.dk/kvaeg/diverse/principles.pdf> Accessed July 5, 2006.
- Danish Cattle. 2006. Genetisk udvikling. <http://www.lr.dk/applikationer/kate/viskategori.asp?ID=ka00400008000010000500> Accessed July 5, 2006.
- Druet, T., J. Soelkner, and N. Gengler. 1999. Use of multi-trait evaluation procedure to improve reliability of early prediction of survival. Online. Available: <http://jds.fass.org/content/vol82/issue9/> *J. Dairy Sci.* 82:2054.
- Ducrocq, V. 1994. Multi trait prediction: Principles and problems. *Proc. 5th World Congr. Genet. Appl. Livest. Prod.* Guelph, Canada, 18:455-462.
- Ducrocq, V., and B. Besbes. 1993. Solution of multiple trait animal models with missing data on some traits. *J. Anim. Breed. Genet.* 110:81-92.
- Ducrocq, V., D. Boichard, A. Barbat, and H. Larroque. 2001. Implementation of an approximate multi-trait BLUP to combine production traits and functional traits into a total merit index. In *Book of Abstract. 52nd Annu. Mtg. EAAP.* Budapest, Hungary.
- Ducrocq, V., and H. Chapuis. 1997. Generalizing the use of the canonical transformation for the solution of multivariate mixed model equations. *Genet. Sel. Evol.* 29:205-224.
- Ducrocq, V., I. Delauney, D. Boichard, and S. Mattalia. 2003. A general approach for international genetic evaluations robust to inconsistencies of genetic trends in national evaluations. *Interbull Workshop, Beltsville, MD.* *Interbull Bull.* 30:101-111.
- Haile-Mariam, M., P. J. Bowman, and M. E. Goodard. 2004. Genetic parameters for fertility, and their correlation with production, type, workability, liveweight, survival index, and cell count. *Aust. J. Agric. Res.* 55:77-88.
- Henderson, C. R. 1975. Best linear unbiased estimation and prediction under a selection model. *Biometrics* 31:423-447.
- Lassen, J., M. Hansen, M. K. Sørensen, G. P. Aamand, L. G. Christensen, and P. Madsen. 2003. Genetic relationship between body condition score, dairy form, mastitis and diseases other than mastitis in first-parity Danish Holstein cows. *J. Dairy Sci.* 86:3730-3735.
- Lindhé, B., and J. Phillipson. 1998. Genetic correlations between production with disease resistance and fertility in dairy cattle and consequences for total merit selection. *Acta Agric. Scand. A Anim. Sci.* 48:216-221.
- Liu, Z., F. Reinhardt, and R. Reents. 2004. A multi-trait MACE model for international bull comparison based on daughter yield deviations. *Proc. Interbull Mtg., Sousse, Tunisia, Interbull Bull.* 32:46-52.
- Madsen, P., and J. Jensen. 2005. A user's guide to DMU. A package for analyzing multivariate mixed models. Version 6, release 4.5, Damarks Jordbrugs Forskning, Foulum, Denmark. <http://dmu.agrsci.dk> Accessed July 5, 2006.
- Miglior, F., B. L. Muir, and B. J. Van Doormaal. 2005. Selection indices in Holstein cattle of various countries. *J. Dairy Sci.* 88:1255-1263.
- Mrode, R. A. 2005. Linear models for the prediction of animal breeding values. CAB International, Wallingford, UK.
- Rupp, R., and D. Boichard. 1999. Genetic parameters for clinical mastitis, somatic cell score, production, udder type traits, and milking ease in first lactation Holsteins. *J. Dairy Sci.* 82:2198-2204.
- Schaeffer, L. 1994. Multiple country comparison of dairy sires. *J. Dairy Sci.* 77:2671-2678.
- Schaeffer, L. R., F. S. Schenkel, and L. A. Fries. 1998. Selection bias on animal model evaluation. *Proc. 6th World Congr. Genet. Appl. Livest. Prod.* Armidale, Australia, 25:501-508.
- Short, T. H., and T. J. Lawlor. 1992. Genetic parameters of conformation traits, milk yield, and herd life in Holsteins. *J. Dairy Sci.* 75:1987-1998.
- Sullivan, P. G., and J. W. Wilton. 2001. Multiple-trait MACE with a variable number of traits per country. *Proc. Interbull Mtg., Budapest, Hungary, Interbull Bull.* 27:68-72.
- Sullivan, P. G., J. W. Wilton, L. R. Schaeffer, G. J. Jansen, J. A. B. Robinson, and O. B. Allen. 2005. Genetic evaluation strategies for multiple traits and countries. *Livest. Prod. Sci.* 92:195-205.
- Sørensen, M. K., P. Berg, J. Jensen, and L. G. Christensen. 1999. Stochastic simulation of breeding schemes for total merit in dairy cattle. Genetic Improvement of Functional Traits in Cattle. Wageningen, the Netherlands, *Interbull Bull.* 23:183-192.
- Sørensen, M. K., J. Jensen, and L. G. Christensen. 2000. Udder conformation and mastitis resistance in Danish first-lactation cows: Heritabilities, genetic and environmental correlations. *Acta Agric. Scand. A Anim. Sci.* 50:72-82.
- Tarrés, J., J. Piedrafit, and V. Ducrocq. 2006a. Validation of an approximate approach to compute genetic correlations between longevity and linear traits. *Genet. Sel. Evol.* 38:65-85.
- Tarrés, J., Z. Liu, V. Ducrocq, F. Reinhardt, and R. Reents. 2006b. Parameter estimation and genetic evaluation of milk production traits from France and Germany with a multi-trait MACE model. *Proc. Interbull Mtg., Kuopio, Finland, Interbull Bull.* 35:76-86.
- Thompson, R., and K. Meyer. 1986. A review of theoretical aspects in the estimation of breeding values for multi-trait selection. *Livest. Prod. Sci.* 15:299-313.
- Van der Werf, J. H. J., J. A. M. van Arendonk, and A. G. De Vries. 1992. Improving selection of pigs using correlated characters. 43rd EAAP Annu. Mtg., Madrid, Spain. Genetic Improvement of Functional Traits in Cattle.
- Villanueva, B., N. R. Wray, and R. Thompson. 1993. Prediction of asymptotic rates of response from selection on multiple traits using univariate and multivariate best linear unbiased predictors. *Anim. Prod.* 57:1-13.