1) Institut für Tierzucht und Tierhaltung der Christian-Albrechts-Universität zu Kiel, Germany
2) Forschungsinstitut für die Biologie landwirtschaftlicher Nutztiere, Dummerstorf, Germany

BIRTE HARDER[1], JÖRN BENNEWITZ[1], NORBERT REINSCH[2],
MANFRED MAYER[2], and ERNST KALM[1]

Effect of missing sire information on genetic evaluation

Dedicated to Prof. Dr. Detlef Leonhard Simon on the occasion of his 75th birthday

Abstract

Stochastic simulation was used to analyse the effect of missing sire information (MSI) on different parameters of genetic evaluation. Eighty proven bulls producing 100 progeny and 40 test bulls producing 50 progeny were simulated. The proportion of MSI was varied in four steps from 10 to 40%. Analyses were carried out for $h^2 = 0.10$ and $h^2 = 0.25$. A sire model was used for simulation and evaluation. The variance of DYD increased with increasing proportion of MSI. Variances of sire breeding values as well as rank correlations decreased with increasing proportion of MSI. The probability that the simulated Top5 and Top10 bulls were placed under the estimated Top5 and Top10 bulls, respectively, decreased with increasing percentage of MSI. The same holds true for the probability of ranking a bull with 10 to 40% MSI under the 5% best bulls with complete pedigree. The loss of response to selection increased up to 8.6% for proven bulls and up to 12.6% for test bulls.

Key Words: breeding value estimation, progeny testing, missing sire information, variance of DYD, genetic gain, bulls

Zusammenfassung

Titel der Arbeit: **Einfluss fehlender Vaterinformationen auf die Zuchtwertschätzung**
Anhand einer stochastischen Simulation wurde der Einfluß fehlender Vaterinformationen auf die Zuchtwertschätzung analysiert. In jeder Wiederholung wurden 80 Altbullen, mit jeweils 100 Nachkommen, und 40 Testbullen mit jeweils 50 Nachkommen simuliert. Der Anteil fehlender Vaterinformationen variierte in 4 Stufen von 10 bis 40%, die für $h^2 = 0,25$ und $h^2 = 0,10$ analysiert wurden. Für die Auswertung wurde ein Vatermodell verwendet. Die Varianzen der DYD stiegen mit zunehmendem Anteil fehlender Abstammung an, wohingegen die Varianzen der Zuchtwerte und die Rangkorrelationen mit zunehmendem Anteil fehlender Abstammung zurückgingen. Dies galt ebenfalls für die Wahrscheinlichkeit, dass die fünf und zehn Bullen mit den besten wahren Zuchtwerten jeder Wiederholung sich auch unter den Bullen mit den besten fünf beziehungsweise zehn Bullen mit den höchsten geschätzten Zuchtwerten wiederfanden. Des weiteren führte die zunehmende fehlende Abstammung auch bei der Wahrscheinlichkeit, Bullen mit 10-40% fehlender Abstammung unter den besten 5% Bullen mit vollständiger Abstammung zu plazieren. Im Zuchtfortschritt war ein Rückgang von bis zu 8.6% für Altbullen und bis zu 12,6% für Testbullen zu verzeichnen.

Schlüsselwörter: Zuchtwertschätzung, Nachkommenschaftsprüfung, fehlende Vaterinformationen, Varianz der DYD, Zuchtfortschritt, Bullen

Introduction

Two kinds of pedigree errors can influence the results of breeding value estimation - incorrect pedigree information (i.e. wrong parentage) and missing pedigree information (i.e. parentage unknown). The fraction of wrong paternity can be substantial. For example, it was estimated to be between 5 and 15% in Denmark (CHRISTENSEN et al., 1982), between 3 and 5% in Israel (RON et al., 1996), around 12% in the Netherlands (BOVENHUIS and VAN ARENDONK, 1991), between 4 and

23% in Germany (GELDERMANN et al., 1986), and around 10% in the UK (VISSCHER et al., 2002). Errors, particularly in sire identification, are a source of bias in the estimation of genetic parameters, of breeding values of individual animals, and of genetic progress (VAN VLECK 1970a, 1970b; ISRAEL and WELLER, 2000; BANOS et al., 2001). According to GELDERMANN et al. (1986), the genetic gain is around 9% (17%) reduced for a trait with a heritability of 0.5 (0.2) and a misidentification rate of 15% compared to a correct pedigree.

Until now only very few studies investigated the amount and the consequences of the second kind of pedigree errors, missing paternity information (SCHENKEL and SCHAEFFER, 2000; ROUGHSEDGE et al., 2001; SCHENKEL et al., 2002). Often, the studies solely investigated the impact of missing pedigree information on estimates of inbreeding (WOLLIAMS and MÄNTYSAARI, 1995; LUTAAYA et al., 1999). The consequence of incomplete pedigree information is an underestimation of inbreeding. According to CASSELL et al. (2003) the frequency of missing information in pedigrees of registered Holsteins and Jerseys in the United States was 3% and 11%, respectively. For the purpose of genetic evaluation this can partly be overcome by genetic grouping (PIERAMATI and VAN VLECK, 1993).

In Germany the percentages of missing pedigree information vary between almost complete pedigree information and 33% missing pedigree information depending on the region (REINHARDT, 2004). In regions with high percentages of missing pedigree information a lower variance of Daughter Yield Deviation (DYD) was observed (BÜNGER, unpublished). This suggests that missing pedigree information could cause less variation in DYD, and therefore influence breeding value estimation.

The aim of this study was to evaluate whether the proportion of missing sire information influences the variance of Yield Deviation (YD) and DYD. Furthermore, the impact of incomplete pedigrees on the variance of breeding values, different probabilities of success, and on loss of response to selection was analysed.

Materials and Methods

Simulation Model

A Monte Carlo simulation program was developed that modelled a two generation dairy cattle population using the following sire model:

$$Y_{ijk} = h_i + s_j + e_{ijk}$$

where Y_{ijk} is the phenotypic record of daughter k with herd i and sire j, h_i is the effect of the ith herd, s_j is the effect of sire j, and e_{ijk} is the residual. The herd effect was sampled from N (0,1), the sire effect from N (0, 0.25h²), and the residual effect from N (0, 1-0.25h²). The heritability was either 0.10 or 0.25 (h² = 0.10, 0.25).

The base generation consisted of 80 proven bulls and 40 test bulls. The second generation was built by the daughters of the proven bulls and of the test bulls. The average number of daughters for each proven bull was 100 and for each test bull 50. The progenies were randomly distributed over 1000 herds. The number of progenies per herd was 10. For each configuration 500 replicates were generated.

Simulation of missing sire information

In a second step a defined proportion of paternity was set to missing. The proportion of missing sire information (MSI) was varied in four steps (MSI = 0.10, 0.20, 0.30, 0.40),

regardless whether the sire was a proven or a test bull. For individuals with MSI phantom parents were created, with all such parents considered to be unrelated. This implies that the production records of daughters with MSI remain in the data sets. Consequently, for each simulated configuration five data sets were generated, one with the true paternity, and the remaining four with the proportion MSI of missing paternity as described above.

Evaluation

The data sets were analysed by the sire model shown above. In the analysis the herd effects were treated as fixed effect, and the sire effect as random. Breeding values (BV) for the sire and herd effects were estimated by BLUP using PEST (GROENEVELD et al., 1998). Variance components were obtained using VCE4 (GROENEVELD et al., 1990) The DYD of a bull is defined as the average of daughters' performance adjusted for fixed and non-genetic random effects of the daughters and genetic effects of the daughters dams (VANRADEN and WIGGANS, 1991). Yield deviations were obtained as:

$$\hat{YD}_{ijk} = y_{ijk} - \hat{h}_j$$

The covariance matrix of the DYD in a sire model is as follows:

$$\text{Var}\left[\text{DYD}\right] = A_s \cdot \sigma_s^2 + K \cdot \sigma_e^2$$

where A_s is the numerator relationship matrix of the sires, K is a matrix with the reciprocal value of the number of daughters per sire on the diagonal and zeros on the off-diagonal, and σ_s^2 and σ_e^2 are the between sire variance (equal to ¼ of the additive genetic variance) and the residual variance, respectively. The following equation was used to estimate the variance of estimated DYD:

$$\text{Var}\left[\hat{DYD}\right] = I \cdot \hat{\sigma}_s^2 + K \cdot \hat{\sigma}_e^2$$

where I is an identity matrix, because all covariances are zero. The result is a column vector containing the variances of the DYD. The variance components were estimated for each replicate using the program VCE 4.0, version 4.2.5 (GROENEVELD, 1998). The variance of DYD within a genetic configuration was calculated as the average value over all replicates.

The variances of estimated sire breeding values were estimated using the following formula:

$$\text{Var}\left[\hat{a}\right] = R \cdot \hat{\sigma}_s^2$$

with $\hat{a}$ being a column vector of the dimension number of sires, R being a matrix with reliabilities (R) of sire breeding values, obtained from PEST (GROENEVELD et al., 1990), on the diagonal and zeros on the off-diagonal, and σ_s^2 being the sire variance.

Rank correlations between simulated breeding values (BV) and estimated breeding values (EBV) were estimated as Spearman's correlation coefficients, using the CORR procedure of the SAS Package (SAS, 1999).

The EBV and the BV of proven bulls and test bulls were ranked for each data set within each configuration and replicate. The probability that the estimated Top5 (with different MSI) were ranked under the simulated Top5 (i.e. the true Top5) was estimated. The same was done for the Top10 proven bulls and test bulls. Similarly, the probability that the estimated Top5 with a MSI > 0 were ranked under the estimated Top5 with a MSI = 0 were calculated. Again, this was also performed for the Top10.

Mean square errors (MSE) were estimated as the average squared deviation of predicted breeding values, predicted herd effects, and predicted YD, respectively, from their corresponding true values. MSE for sire effects, herd effects, and YD were estimated within each configuration as the average value over all replicates.

Following Mrode (1996) the genetic gain (G) can be expressed as:

$$\Delta G = i \cdot R^{0.5} \cdot \sigma_a,$$

where i is the selection intensity, R is the reliability and σ_a is the additive genetic standard deviation. In a progeny testing scheme the reliability of a sire is approximated by $R = \dfrac{n}{n + \lambda}$, with $\lambda = \dfrac{4 - h^2}{h^2}$. Following this, R is affected by MSI. Under the assumption that i and σ_a are not affected by MSI, the efficiency of a breeding scheme with regard to response to selection is a function of the MSI. The relative efficiency (relative to complete sire information) can be expressed as:

$$\text{Efficiency} = \left[\frac{\left(n_{MSI} / \left(n_{MSI} + \lambda \right) \right)}{\left(n_0 / \left(n_0 + \lambda \right) \right)} \right]^{0.5}$$

where n_0 is the number of daughters with complete sire information, and n_{MSI} is the number of daughters for the different proportions of MSI (MSI = 0.10, 0.20, 0.30, 0.40). The relative loss in response is, Loss = 1-Efficiency.

Table 1

Average estimates of sire variance (σ_s^2) and residual variance (σ_e^2) over all replicates as a function of missing sire information (MSI) for proven bulls (PB) and for test bulls (TB) for two levels of heritability (h² = 0.25 and h² = 0.10) (Durchschnittliche Vater- und Restvarianzen über alle Wiederholungen in Abhängigkeit vom Anteil fehlender Abstammung dargestellt für Alt- und Testbullen und für zwei Heritabilitätsniveaus (h² = 0,25 und h² = 0,10))

	VC	percentage of missing sire information (MSI)				
		0%	10%	20%	30%	40%
PB, h² = 0.25	σ_s^2	0.0620	0.0620	0.0620	0.0622	0.0623
	σ_e^2	0.937	0.938	0.938	0.937	0.937
PB, h² = 0.10	σ_s^2	0.0246	0.0246	0.0246	0.0247	0.0247
	σ_e^2	0.975	0.976	0.975	0.974	0.975
TB, h² = 0.25	σ_s^2	0.0610	0.0611	0.0612	0.0608	0.0606
	σ_e^2	0.937	0.938	0.938	0.937	0.937
TB, h² = 0.10	σ_s^2	0.0245	0.0245	0.0245	0.0242	0.0240
	σ_e^2	0.975	0.976	0.975	0.974	0.975

Results

The results for the proven bulls and test bulls are shown separately.

Daughter yield deviation

The between sire variance component as well as the residual variance component were not influenced by the proportion of MSI (Table 1). The impact of MSI on the variance of the DYD both for proven bulls and for test bulls is shown in Table 2. The variance increased by about 8% for proven bulls, and by about 14% for test bulls at a heritability of 0.25, comparing 0% and 40% MSI. At a heritability of 0.10, the impact of MSI was higher. The variance increased by about 18% for the proven bulls and by about 28% for the test bulls comparing complete and 40% missing sire information. For both heritabilities the absolute values of the DYD were higher for test bulls than for proven bulls.

Table 2
Average variance of the Daughter yield deviation (DYD) over all replicates as a function of missing sire information (MSI) for proven bulls (PB) and for test bulls (TB) for two levels of heritability (h² = 0.25 and h² = 0.10) (Durchschnittliche Varianz der Daughter Yield Deviation (DYD) über alle Wiederholungen in Abhängigkeit vom Anteil fehlender Abstammung dargestellt für Alt- und Testbullen und für zwei Heritabilitätsniveaus (h² = 0,25 und h² = 0,10))

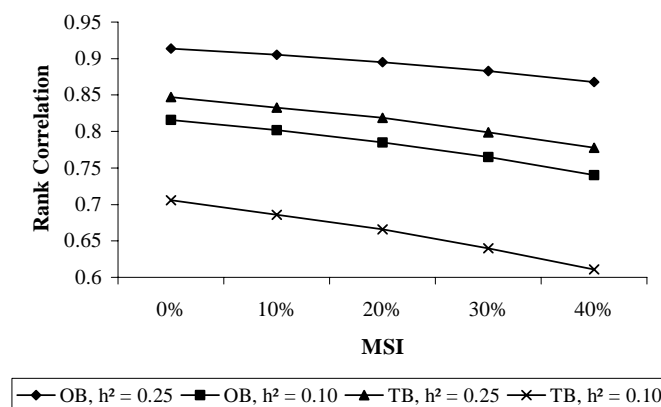|	percentage of missing sire information (MSI)				
	0%	10%	20%	30%	40%
PB, $h^2 = 0.25$	0.0706	0.0715	0.0726	0.0744	0.0764
PB, $h^2 = 0.10$	0.0335	0.0344	0.0356	0.0374	0.0396
TB, $h^2 = 0.25$	0.0784	0.0802	0.0827	0.0856	0.0897
TB, $h^2 = 0.10$	0.0424	0.0444	0.0469	0.0500	0.0543

Fig. 1: Average value of Spearman's rank correlations over all replicates between true and estimated breeding values as a function of missing sire information (MSI) for proven bulls (PB) and test bulls (TB) and for two levels of heritability (h²) (Durchschnittliche Spearmannsche Rangkorrelationen zwischen wahren und geschätzten Zuchtwerten über alle Wiederholungen in Abhängigkeit vom Anteil fehlender Abstammung dargestellt für Alt- und Testbullen und für zwei Heritabilitätsniveaus (h²))

Rank Correlations

Rank correlations between EBV and BV are presented in Figure 1. With increasing MSI rank correlations decreased for all four data sets. Rank correlations for proven

bulls ranged from 0.91 with complete sire information to 0.87 with 40% MSI, and from 0.82 with complete sire information to 0.74 with 40% MSI, for heritabilities of 0.25 and 0.10, respectively. Rank correlations for test bulls ranged from 0.85 with complete sire information to 0.78 with 40% MSI, and from 0.71 with complete sire information to 0.61 with 40% MSI, for heritabilities of 0.25 and 0.10, respectively. The decrease of rank correlations with increasing MSI was highest for test bulls at heritability of 0.10. In this case, the increase of MSI up to 40% led to 10% lower rank correlations compared to full sire information.

Table 3
Average variances of estimated breeding values (EBV) and true breeding values (BV) over all replicates as a function of missing sire information (MSI) for proven bulls (PB) and test bulls (TB), and for two different levels of heritability ($h^2 = 0.25$ and $h^2 = 0.10$) (Durchschnittliche Varianz der geschätzten Zuchtwerte (EBV) und der wahren Zuchtwerte (BV) über alle Wiederholungen in Abhängigkeit vom Anteil fehlender Abstammung dargestellt für Alt- und Testbullen und für zwei Heritabilitätsniveaus ($h^2 = 0,25$ und $h^2 = 0,10$))

	BV	percentage of missing sire information (MSI)				
		0%	10%	20%	30%	40%
PB, $h^2 = 0.25$	0.0627	0.0530	0.0524	0.0514	0.0502	0.0486
TB, $h^2 = 0.25$	0.0612	0.0463	0.0452	0.0437	0.0419	0.0397
PB, $h^2 = 0.10$	0.0250	0.0172	0.0167	0.0161	0.0153	0.0144
TB, $h^2 = 0.10$	0.0240	0.0132	0.0126	0.0119	0.0110	0.0101

Variances of breeding values
According to Table 3, the variance of the true breeding values for proven bulls was 0.0627 at a heritability of 0.25, and was 0.025 at a heritability of 0.10. For test bulls, the variance of the true breeding values was 0.0612 at a heritability of 0.25, and was 0.024 at a heritability of 0.10. These values were close to the assumed simulation parameters of 0.625 and 0.025, respectively. With complete sire information, the estimated variance of sire breeding values for proven bulls was about 16% lower at a heritability of 0.25, and about 31% lower at a heritability of 0.10 compared to the variance of the simulated breeding values. The variance of proven bull's breeding values decreased by about 8% at heritability of 0.25, and by about 16% at heritability of 0.10, comparing complete and 40% MSI.
For test bulls, the variance of sire breeding values with complete pedigree was about 24% lower at a heritability of 0.25, and about 45% lower at a heritability of 0.10 compared to the corresponding variance of the true breeding values. At a level of 40% MSI, the variance of test bull's breeding values decreased by 14% at a heritability of 0.25, and by 24% at a heritability of 0.10 compared to full pedigree information. A comparison of the different configurations shows that the decline of the variance of sire breeding values was stronger for test bulls and for the lower heritability.
In Figure 2 the reliabilities as a function of the proportion of missing sire information are presented. The reliability was highest for proven bulls at $h^2 = 0.25$ and lowest for test bulls at $h^2 = 0.10$. For all four configurations the reliability declined with increasing proportion of MSI. For proven bulls the reliabilities ranged between 0.85 and 0.78 at $h^2 = 0.25$ and between 0.69 and 0.58 at $h^2 = 0.10$ at the different levels of MSI. For test bulls the reliabilities ranged between 0.74 and 0.64 at $h^2 = 0.25$ and between 0.53 and 0.40 at $h^2 = 0.10$ at the different levels of MSI.
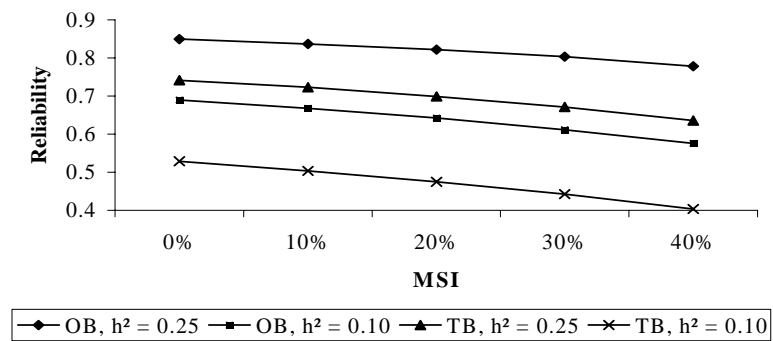
Fig. 2: Average reliabilities over all replicates as a function of missing sire information (MSI) for proven bulls (PB) and test bulls (TB) and for two levels of heritabilities (h²) (Durchschnittliche Sicherheiten über alle Wiederholungen in Abhängigkeit vom Anteil fehlender Abstammung dargestellt für Alt- und Testbullen und für zwei Heritabilitätsniveaus (h²))

Probability of success

The average probability that the estimated Top5 proven bulls with complete sire information were the simulated Top5 proven bulls was 0.65 and 0.52 for heritabilities of 0.25 and 0.10, respectively (Figure 3). The increase of the proportion of missing sire information up to 40% caused a steady decline of the probability of success down to 0.58 and 0.44 for h² = 0.25 and h² = 0.10, respectively.
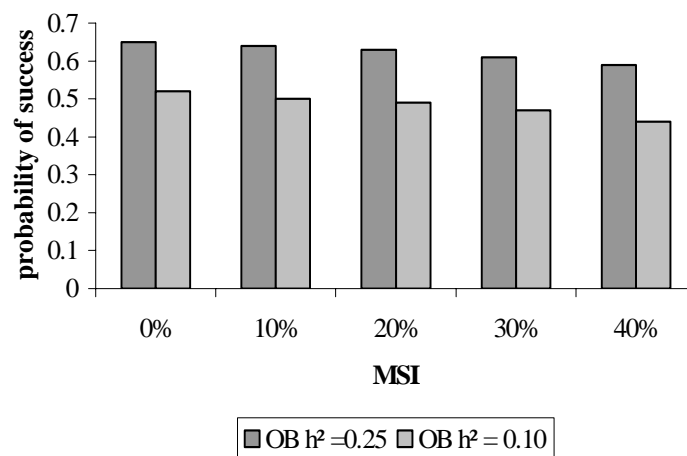

Fig. 3: Average probability that the estimated Top5 were placed under the true Top5 over all replicates as a function of missing sire information (MSI) for proven bulls (PB) at heritability (h²) of 0.25 and of 0.10 (Durchschnittliche Wahrscheinlichkeit über alle Wiederholungen, dass sich die geschätzten TOP5 Bullen unter den wahren Top5 Bullen plazieren, in Abhängigkeit vom Anteil fehlender Abstammung, dargestellt für Altbullen und Heritabilitäten von 0,25 und 0,10)

The probability of success for test bulls was nearly on the same level as for proven bulls at both heritabilities (Figure 4). It decreased by 11% at h² = 0.25 and by 15% at h² = 0.10 comparing full and 40% MSI.

The level of probabilities of success for the Top10 was about 10 percentage points higher than for the Top5 for proven bulls and test bulls at both heritabilities (not shown). Considering the impact of MSI on the probability of success the results of the Top10 alternative were similar to those of the Top5 alternative.
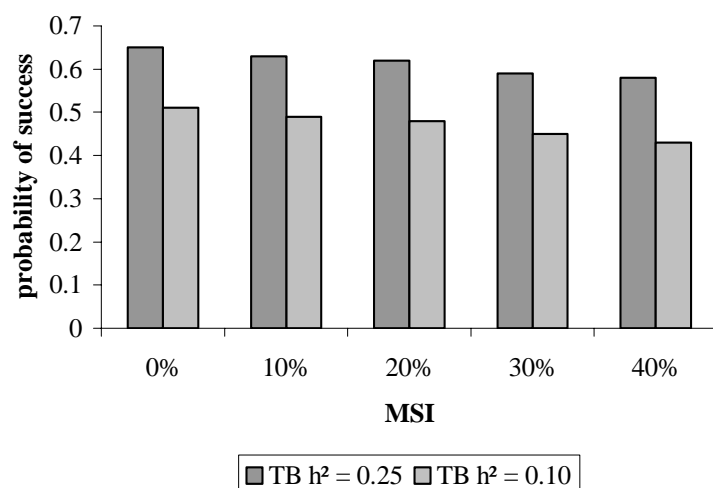
Fig. 4: Average probability that the estimated Top5 were placed under the true Top5 over all replicates as a function of missing sire information (MSI) for the test bulls (TB) at heritability of 0.25 and 0.10 (Durchschnittliche Wahrscheinlichkeit über alle Wiederholungen, dass sich die geschätzten TOP5 Bullen unter den wahren Top5 Bullen plazieren, in Abhängigkeit vom Anteil fehlender Abstammung, dargestellt für Testbullen und Heritabilitäten von 0,25 und 0,10)

Table 4
Average probability of ranking a bull with 10 to 40% missing sire information (MSI) under the 5% best of the 0% alternative over all replicates for proven bulls (PB) and test bulls (TB), and for two different levels of heritability ($h^2 = 0.25$ and $h^2 = 0.10$) (Durchschnittliche Wahrscheinlichkeit einen Bullen bei 10-40% fehlender Abstammung unter den 5% besten Bullen bei kompletter Abstammung zu plazieren, über alle Wiederholungen in Abhängigkeit vom Anteil fehlender Abstammung dargestellt für Alt- und Testbullen und für zwei Heritabilitätsniveaus ($h^2 = 0,25$ und $h^2 = 0,10$))

	percentage of missing sire information (MSI)			
	10%	20%	30%	40%
PB, $h^2 = 0.25$	0.85	0.82	0.80	0.75
PB, $h^2 = 0.10$	0.82	0.77	0.72	0.62
TB, $h^2 = 0.25$	0.72	0.70	0.64	0.57
TB, $h^2 = 0.10$	0.70	0.62	0.53	0.42

A comparison between data sets with incomplete sire information (10, 20, 30, 40% MSI) and complete sire information is shown in Table 4. The probability of ranking a bull with 10 to 40% MSI under the 5% best bulls with complete sire information was evaluated. On average over all replicates 75 to 85% at $h^2 = 0.25$, and 62 to 82% at $h^2 = 0.10$ of the proven bulls in the different alternatives of MSI were ranked under the 5% best of the 0% alternative. For test bulls the average over all replicates was 57 to 72%, and 42 to 70% for $h^2 = 0.25$ and $h^2 = 0.10$, respectively. The decrease of probability of success was strongest for test bulls with low heritability (Table 4). In this case the probability of success decreased by about 28 percentage points comparing 10% and 40% MSI.

Mean Square Error (MSE)
Tables 5 and 6 show that the MSE of sire and herd effects increased with increasing proportion of missing pedigree information for proven and for test bulls at both

heritabilities. The level of MSE was higher for test bulls and for heritability of 0.25. A systematic over- or underestimation of sire breeding values and of herd effects could not be observed (not shown).

Table 5
Average Mean Square Errors (MSE) of estimated sire breeding values over all replicates as a function of missing sire information (MSI) for proven bulls (PB) and test bulls (TB), and for two different levels of heritability ($h^2 = 0.25$ and $h^2 = 0.10$) (Durchschnittliche Mean Square Errors (MSE) der geschätzten Zuchtwerte über alle Wiederholungen in Abhängigkeit vom Anteil fehlender Abstammung dargestellt für Alt- und Testbullen und für zwei Heritabilitätsniveaus ($h^2 = 0,25$ und $h^2 = 0,10$))

	percentage of missing sire information (MSI)				
	0%	10%	20%	30%	40%
PB, $h^2 = 0.25$	0.0094	0.0103	0.0114	0.0126	0.0143
PB, $h^2 = 0.10$	0.0078	0.0084	0.0090	0.0098	0.0107
TB, $h^2 = 0.25$	0.0162	0.0176	0.0190	0.0208	0.0229
TB, $h^2 = 0.10$	0.0117	0.0124	0.0131	0.0139	0.0148

Table 6
Average Mean Square Errors (MSE) of estimated herd effects over all replicates as a function of missing sire information (MSI) for two different levels of heritability ($h^2 = 0.25$ and $h^2 = 0.10$) (Durchschnittliche Mean Square Errors (MSE) der geschätzten Betriebseffekte über alle Wiederholungen in Abhängigkeit vom Anteil fehlender Abstammung dargestellt für zwei Heritabilitätsniveaus ($h^2 = 0,25$ und $h^2 = 0,10$))

	percentage of missing sire information (MSI)				
	0%	10%	20%	30%	40%
$h^2 = 0.25$	0.0954	0.0960	0.0966	0.0972	0.0977
$h^2 = 0.10$	0.0987	0.0989	0.0990	0.0991	0.0995

Table 7 presents the MSE of the YD. The MSE for the YD increased with increasing proportion of MSI for proven bulls and for test bulls at both heritabilities. For proven bulls the MSE of the YD ranged between 0.010 and 0.016 at $h^2 = 0.25$ and between 0.010 and 0.017 at $h^2 = 0.10$. For test bulls the MSE of the YD ranged between 0.010 and 0.13 at $h^2 = 0.25$ and between 0.010 and 0.13 at $h^2 = 0.10$.

Table 7
Average Mean Square Errors (MSE) of estimated YD over all replicates as a function of missing sire information (MSI) for proven bulls (PB) and test bulls (TB), and for two different levels of heritability ($h^2 = 0.25$ and $h^2 = 0.10$) (Durchschnittliche Mean Square Errors der geschätzten YD über alle Wiederholungen in Abhängigkeit vom Anteil fehlender Abstammung dargestellt für Alt- und Testbullen und für zwei Heritabilitätsniveaus ($h^2 = 0,25$ und $h^2 = 0,10$))

	percentage of missing sire information (MSI)				
	0%	10%	20%	30%	40%
PB, $h^2 = 0.25$	0.095	0.112	0.129	0.146	0.162
PB, $h^2 = 0.10$	0.099	0.116	0.133	0.150	0.167
TB, $h^2 = 0.25$	0.095	0.104	0.113	0.122	0.130
TB, $h^2 = 0.10$	0.099	0.107	0.116	0.125	0.134

Losses in response to selection

The impact of missing sire information on the loss in response to selection is shown in Table 8. The loss in response to selection increased with increasing MSI for proven and for test bulls at both heritabilities. For proven bulls the losses in response to selection ranged between 0.6% and 4% at h² = 0.25 and between 1.5% and 8.6% at h² = 0.10 at the different levels of MSI. For test bulls the losses in response to selection ranged between 1.3% and 7.4% at h² = 0.25 and between 2.4% and 12.6% at h² = 0.10 at the different levels of MSI (relative to 0% MSI).

Table 8
Average losses of selection response (%) over all replicates as a function of missing sire information (MSI) for proven bulls (PB) and test bulls (TB), and for two different levels of heritability (h² = 0.25 and h² = 0.10) (Durchschnittlicher Rückgang des Selektionserfolges (%) über alle Wiederholungen in Abhängigkeit vom Anteil fehlender Abstammung dargestellt für Alt- und Testbullen und für zwei Heritabilitätsniveaus (h² = 0,25 und h² = 0,10))

| | percentage of missing sire information (MSI) | | | |
	10%	20%	30%	40%
PB, h² = 0.25	0.63	1.53	2.73	4.24
PB, h² = 0.10	1.55	3.50	5.77	8.64
TB, h² = 0.25	1.27	2.88	4.88	7.39
TB, h² = 0.10	2.43	5.28	8.61	12.57

Discussion

Simulation study

Until now only very few studies dealing with the impact of missing pedigree information on estimation of genetic parameters and breeding values have been carried out (SCHENKEL and SCHAEFFER, 2000; ROUGHSEDGE et al. 2000; SCHENKEL et al., 2001). Because of the fact that in practise often only sire information is unknown, the focus of the present study lay on the impact of missing paternity on variances of YD, DYD, sire breeding values, and on probabilities of success. In contrast to this, SCHENKEL and SCHAEFFER (2000) analysed missing pedigree information by deleting both sire and dam information, and LUTAAYA et al. (1999) evaluated the impact of missing dams on average inbreeding coefficients. In the present study the proportion of missing sire information was increased up to 40%, following the figures of missing paternity in Germany. This is in accordance with LUTAAYA et al. (1999). The authors increased the proportion of missing dams by up to 50%.

Yield deviation, Daughter yield deviation

The increase of the MSE of the YD with increasing proportion of MSI is caused by the incline of the MSE of the herd effects (Table 6). This is due to the fact that with increasing proportion of MSI a proportion of halfsib cows are treated as unrelated. Therefore, the variance-covariance-matrix of the observation is estimated incorrectly for data sets with missing sire information.

The results clearly show the impact of the amount of sire information on the variance of the DYD. According to the formula the variance of the DYD is influenced by the between sire variance component, by the residual variance component and by the number of daughters per sire. The variance component estimation showed that neither of the variance components were influenced by the proportion of missing sire information (Table 1). In contrast to this, the number of daughters within sire decreased with increasing percentage of MSI and therefore, the diagonal elements of the matrix **K** increased with inclining MSI. This fact could explain the increasing variance of DYD with increasing MSI.

In accordance with this, the lower number of daughters per test bull caused the higher variation of DYD for test bulls compared to proven bulls.

Rank Correlations

It was shown that missing sire information led to decreasing rank correlations (Figure 1). This was also pointed out by SCHENKEL et al. (2000). The authors described a decrease by 5 to 10% regarding full and 15% missing pedigree information. The decreasing rank correlations between true and estimated breeding values (Figure 1) could be explained by the MSE of sire breeding values, which increased with higher proportions of missing sire information for proven bulls and even more for test bulls (Table 5). This is in accordance with the stronger decline of test bulls rank correlations.

Variances in breeding values

The fact that the simulated breeding values were close to the simulation parameters at both heritabilities and for proven bulls and test bulls supported the validity of the simulation and analysis model. The results indicate that missing paternity shrinks the variance of breeding values both for proven and for test bulls. Similar results were shown for incorrect identification of sires (ISRAEL and WELLER, 2000; VAN VLECK, 1970a). ISRAEL and WELLER (2000) pointed out, that pedigree errors reduced the estimated breeding values of elite bulls assumed to be sires of inferior daughters and increase the estimated breeding values of low ranking bulls.

With complete pedigree information, the EBV were lower than the BV for proven bulls and for test bulls, respectively. This could be explained by the MSE of sire breeding values, which already occurred without missing sire information (Table 5).

The decline of variance of sire breeding values with increasing MSI could be explained by the decreasing reliability (Figure 2). The reliability is by definition influenced by the predicted error variance (PEV), which declined with decreasing number of daughters per sire (not shown).

In accordance with this, the decline of the variance of test bull's breeding values with increasing percentage of MSI was stronger than for proven bulls, because of their lower number of daughters. A similar effect was caused by the low heritability, which contributes to the stronger decrease at $h^2 = 0.10$.

Probability of success

For the breeding organisations the impact of missing pedigree information on the ranking of the top bulls, and moreover on the absolute height of their breeding values is most important.

The decreasing probability that the estimated Top5 were placed under the true Top5 bulls was caused by the increasing MSE of estimated sire breeding values.

It is well known that higher heritabilities lead to a more accurate breeding value estimation, therefore the probabilities of success were higher for heritability of 0.25 than for h² = 0.10. For the analysed Top10 the probabilities of success was about 10 percent points higher than for the Top5 alternative. This was due to the higher number of possible successful outcomes for the Top10 alternative.

The comparisons of estimated breeding values at different amounts of MSI showed that especially for young bulls and low heritabilities the impact of MSI on the absolute height of the breeding values was immense. At the low heritability on average over all replicates only 42% of the test bulls had breeding values that were high enough to be ranked under the 5% best, comparing full and 40% missing pedigree information.

Both the comparisons of estimated and simulated breeding values, and the comparisons among the estimated breeding values show that breeding organisations in regions with high percentages of missing pedigree information are disadvantaged.

Losses in response to selection

The results indicate that missing sire information leads to a reduction in response to selection especially for traits with low heritabilities such as functional traits. Similar results were shown by VISSCHER et al. (2002) for incorrect pedigrees. According to VISSCHER et al. (2002) the genetic progress is reduced by 6% for an error rate of 20%, a heritability of 25%, and 50 progeny per sire. The authors pointed out, that their values for losses in response to selection are similar to the predicted increase in genetic gain if including additional traits in a national selection index.

Table 9
Average Mean Square Errors (MSE) of estimated herd effects over all replicates as a function of MSI for two different levels of heritability (h² = 0.25 and h² = 0.10), when daughters with MSI are eliminated from the data sets (Durchschnittliche Mean Square Errors (MSE) der geschätzten Betriebseffekte über alle Wiederholungen in Abhängigkeit vom Anteil fehlender Abstammung dargestellt für zwei Heritabilitätsniveaus (h² = 0,25 und h² = 0,10) unter der Annahme, dass Töchter mit fehlender Vaterinformation vom Datensatz eliminiert werden)

	percentage of missing sire information (MSI)				
	0%	10%	20%	30%	40%
h² = 0.25	0.0954	0.1074	0.1229	0.1441	0.1744
h² = 0.10	0.0987	0.1114	0.1271	0.1446	0.1801

Deleting daughters with MSI from the data sets

In order to evaluate the contribution of having the records of daughters with MSI included in the evaluation, MSE of herd effects were calculated for these data sets, where daughters with MSI were deleted. Table 9 presents the effect of eliminating the daughters with MSI from the data sets on the MSE of the herd effects. The MSE of the herd effects increased by about 82% comparing complete and 40% missing sire information for both heritabilities. The comparison of Table 6 and 9 shows that the impact of deleting the daughters with MSI from the data sets on the estimation of herd effects is much stronger than letting the daughters with MSI in the data sets.

The situation of missing sire information and letting the daughter records remain in the data sets leads to an increase in the MSE of herd effects, because the variance-covariance-matrix of the observation is estimated incorrectly. This is due to the fact

that daughters with MSI are treated as unrelated although they are related. In contrast to this, the variance-covariance-matrix of the observation is estimated correctly if the particular daughter records are eliminated from the data sets, but the amount of data records is reduced with increasing number of MSI. The higher MSE of the herd effects in the situation where daughter records with MSI were eliminated from the data sets show that it is beneficial to have the daughter records included to provide improved estimates of herd effects.

Conclusions

The hypothesis that MSI cause less variance of DYD has to be rejected. But although MSI cause an increase in the variance of the DYD the results show that missing sire information has an enormous impact on variance of sire breeding values, on the different probabilities of success, and on the response to selection. In accordance with this, breeding organisations in regions with high percentages of missing pedigree information are less competitive. To account for the decreasing number of observations per sire with increasing percentage of missing pedigree, the breeding organisations have the possibility to increase the number of daughters per test bull. Admittedly, this implies higher costs for the breeding programs. Another option is to recapture some of the sire information by narrowing down the true sire to a small group of animals and incorporating the probability that each one of them is the true sire into the genetic evaluation. Another possibility is to establish performance testing in test herds, to ensure that complete pedigree information is available.

References

BANOS, G.; WIGGANS, G. R.; POWELL, R. L.:
  Impact of paternity errors in cow identification on genetic evaluations and international comparisons. J. Dairy Sci. **84** (2001), 2523–2529
BOVENHUIS, H.; VAN ARENDONK, J. A. M.:
  Estimation of milk protein gene frequencies in crossbred cattle with maximum likelihood. J. Dairy Sci. **74** (1991), 2728–2736
BÜNGER, A.:
  Unpublished, 2004
CASSELL, B. G.; ADAMEC, V.; PEARSON, R. E.:
  Effect of incomplete pedigrees on estimates of inbreeding and inbreeding depression for days to first service and summit milk yield in Holsteins and Jerseys. J. Dairy Sci. **86** (2003), 2967–2976
CHRISTENSEN, L. G.; MADSEN, P.; PETERSEN, J.:
  The influence of incorrect sire-identification on the estimate of genetic parameters and breeding values. Proc. 2nd World Congr. Genet. Appl. Livest. Prod., Madrid, Spain **7** (1982), 200-208
FALCONER, D. S.; MACKAY, T. F. C.:
  Introduction to Quantitative Genetics (1995) (4th Edition). Addison Wesley Longman, New York
GELDERMANN, H.; PIEPER, U.; WEBER, W. E.:
  Effect of misidentification on the estimation of breeding value and heritability in cattle. J. Anim. Sci. **63** (1986), 1759–1768
GROENEVELD, E.,:
  Pest User's guide. Institute of Animal Husbandry and Animal Sciences, Mariensee, Germany (1990)
GROENEVELD, E.:
  VCE User's guide. Version 4.2. Institute of Animal Husbandry and Animal Sciences, Mariensee, Germany (1998)
ISRAEL, C.; WELLER, J. I.:
  Effect of misidentification on genetic gain and estimation of breeding value in dairy cattle populations. J. Dairy Sci. **83** (2000), 181–187
LUTAAYA, E.; MISZTAL, I.; BERTRAND, J. K.; MARBY, J. W.:
  Inbreeding in populations with incomplete pedigrees. J. Anim. Breed. Genet. **116** (1999), 475–480

MRODE, R. A.:
 Linear Models for the Prediction of Animal Breeding Values, CAB International, Wallington, Oxon, UK, 1996

PIERAMATI, C., VAN VLECK, L. D.:
 Effect of genetic groups on estimates of additive genetic variance. J. Anim. Sci. 71 (1993), 66–70

REINHARDT, F.:
 Personal communication, 2004

RON, M.; BLANC, Y.; BAND, M.; EZRA, E.; WELLER, J.I.:
 Misidentification rate in the Israeli dairy cattle population and its implications for genetic improvement. J. Dairy Sci. 79 (1996), 676–681

ROUGHSEDGE, T.; BROTHERSTONE, S.; VISSCHER, P. M.:
 Bias and Power in the estimation of a maternal family component in the presence of incomplete and incorrect pedigree information. J. Dairy Sci. 84 (2001), 944–950

SAS:
 SAS/STAT User's guide, Version 8, SAS Institute, Cary, NC, USA, 1999

SCHENKEL; F. S.; SCHAEFFER, L. R.:
 Effects of nonrandom parental selection on estimation of variance components. J. Anim. Breed. Genet. 117 (2000), 225–239

SCHENKEL, F. S.; SCHAEFFER, L. R.; BOETTCHER, P. J.:
 Comparison between estimation of breeding values and fixed effects using Bayesian and empirical BLUP estimation under selection on parents and missing pedigree information. Genet. Sel. Evol. 34 (2002), 41–59

WOOLLIAMS, J. A.; MÄNTYSAARI, E. A.:
 Genetic contributions of Finnish Ayrshire bulls over four generations. Animal Science 61 (1995), 177–187

VANRADEN, P. M.; WIGGANS, G. R.:
 Derivation, calculation, and use of national animal model information. J. Dairy Sci. 74 (1991), 2737–2746

VAN VLECK, L. D.:
 Misidentification in estimating the paternal-sib correlation. J. Dairy Sci. 53 (1970a), 1469–1474

VAN VLECK, L. D.:
 Misidentification and sire evaluation. J. Dairy Sci. 53 (1970b), 1697–1702

VISSCHER; P. M.; WOOLLIAMS, J. A.; SMITH, D.; WILLIAMS, J. L.:
 Estimation of pedigree errors in the UK dairy population using microsatellite markers and the impact on selection. J. Dairy Sci. 85 (2002), 2368–2375

Corresponding Author
BIRTE HARDER
Institut für Tierzucht und Tierhaltung, Christian-Albrechts-Universität
Herrmann-Rodewald-Strasse 6
24118 KIEL, GERMANY

E-mail: bharder@tierzucht.uni-kiel.de