

Wrong and Missing Sire Information Affects Genetic Gain in the Angeln Dairy Cattle Population

K. Sanders,¹ J. Bennewitz, and E. Kalm

Institut für Tierzucht und Tierhaltung der Christian-Albrechts-Universität zu Kiel, D-24098 Kiel, Germany

ABSTRACT

In the present study, molecular genetic markers were used to help estimate the degree of wrong sire information in the German Angeln dairy cattle population. Sixteen polymorphic microsatellite markers were genotyped on 5 different paternal half-sib families with a total of 805 daughters. For the genotyping process, blood samples of the daughters and semen samples of the sires were used. Allelic frequencies and exclusion probabilities were estimated. The simultaneous effect of wrong (WSI) and missing sire information (MSI) on the reliability of estimated breeding values and on the genetic gain was investigated using deterministic simulations. For these simulations, different values for the number of daughters per sire, heritability, WSI, and MSI were chosen. The estimated proportion of the WSI was 7% in the German Angeln dairy cattle population. The combined impact of WSI and MSI on the genetic gain was relatively large, especially in the case of small progeny size per sire and low heritability. The impact of WSI was more harmful than MSI on response to selection.

Key words: exclusion probability, genetic response, missing sire information, wrong sire information

INTRODUCTION

Wrong sire information (WSI) is a well-known problem in the estimation of breeding values for dairy cattle. Several studies estimated the proportion of WSI at 3 to 23% in the Holstein Friesian breed (Visscher et al., 2002; Ron et al., 2003; Weller et al., 2004). The presence of WSI reduces realized genetic response (relative to expected response) because the estimates of heritability are biased downward (Israel and Weller, 2000; Banos et al., 2001; Visscher et al., 2002). Additionally, WSI might have a more important effect on genetic gain of lowly heritable traits, because in this case, the impact of pedigree information on the EBV using BLUP is

higher. Christensen et al. (1982) and Weller et al. (2004) mentioned different reasons for paternity errors, which can originate from AI companies, recording by the farmer, or genotyping service, and can arise because of human or technical error.

The second source of pedigree errors is missing sire information (MSI). In contrast to WSI, until now there has been comparatively little information on the extent and impact of this source of error, but as discussed by Harder et al. (2005), the proportion of MSI can be substantial. Harder et al. (2005) pointed out that MSI influences the variance of estimated sire breeding values and reduces the response to selection.

The objective of the present study was to estimate the proportion of wrong sire information in the Angeln dairy cattle population using molecular marker information. Furthermore, the consequences of WSI and MSI on the genetic gain were investigated by deterministic simulation.

MATERIALS AND METHODS

Field Data

The red Angeln breed is a small breed located in the northern part of Germany. Since 1960, different red cattle breeds have been crossed with the Angeln breed. Savaş et al. (1998) reported that 40% of the Angeln population had proportions ranging from 13 to 37% of genetics from other cattle breeds such as Red Holstein, Swedish Red and White, and Finnish Ayrshire. The breed is listed in the database of the European Association of Animal Production (EAAP, 2004), where additional information about it can be found. In the present study, a daughter design was used (Weller et al., 1990) in which 5 paternal half-sib families with a total of 805 daughters were selected. The family size ranged from 123 daughters in family 1 to 199 daughters in family 4; the average was 161 daughters per sire.

Genotyping Process

In 2000, blood samples of the daughters were collected on 41 farms, and semen samples were taken from the 5 sires. No blood samples of the daughters' dams

Received January 7, 2005.

Accepted September 1, 2005.

¹Corresponding author: ksanders@tierzucht.uni-kiel.de

were available. The DNA was extracted using the silica-gel method following Myakishev et al. (1995). The DNA extraction was carried out as follows: 325 μ L of whole blood (containing 50 mM EDTA) or semen was mixed with 650 μ L of bind mix in a 1.5-mL minicentrifuge tube. The combination was incubated in a hybridization oven at 37°C for 15 min (lysis), and the pellet was collected by centrifugation at 5,000 rpm for 10 s in an Eppendorf centrifuge (Eppendorf AG, Hamburg, Germany). The supernatant was poured off. This procedure was conducted 3 times. The pellet was resuspended in 1.0 mL of guanidine solution by vortexing and was incubated in a hybridization oven (37°C). The supernatant was then poured off. The guanidine washing procedure and the incubation were then repeated. The pellet was resuspended in 1 mL of propanol wash, incubated in a hybridization oven, and centrifuged at 5,000 rpm for 10 s. The supernatant was then poured off. This washing procedure was conducted twice. Finally, the pellet was dissolved in 1 mL of ethanol (once) and centrifuged; the supernatant was poured off. The probe was dried for 45 min under vacuum in an exsiccator. The pellet was then resuspended in 400 μ L of Tris-HCl buffer (pH 8.0) and dissolved overnight at 4°C.

The 805 daughters and the 5 sires were genotyped for 16 microsatellite markers. These markers are located on 5 different chromosomes (BTA6, 14, 16, 18, and 27) and were selected from previously published bovine marker maps (USDA cattle genome marker maps: <http://www.marc.usda.gov/genome/genome.html>; INRA BOVMAP database: <http://locus.jouy.inra.fr/>). (The properties of the involved markers will be described in the Results section.) They were chosen because of their highly polymorphic character, and their genotypic information will be used in a subsequent QTL mapping project. The PCR for the microsatellite markers was completed on an MJ Research PTC-200 thermocycler (Global Medical Instrumentation Inc., Ramsey, MN). The electrophoresis of the fluorescent-labeled microsatellite markers was carried out using the MegaBACE 500 Analysis System (Amersham Biosciences Europe GmbH, Freiburg, Germany) and analyzed with the MegaBACE Genetic Profiler Software Suite v2.2 (Amersham Biosciences Europe GmbH). The genotypes were transferred into the ADRDB database (Reinsch, 1999) and checked for agreement with the Mendelian laws of inheritance with the program GEN-CHECK (Bennewitz et al., 2002). To exclude genotyping mistakes, animals involved in conflicts were genotyped for a second time. However, most conflicts (i.e., a violation of the Mendelian laws of inheritance) remained after the second genotyping. Within the set of conflicts that remained unsolved, a paternity was declared

wrong if a conflict between a daughter and its putative sire was observed at ≥ 3 loci.

Marker Characteristics and Estimation of Pedigree Errors

The allelic frequencies were estimated by maximum likelihood with the following log-likelihood function (Brka et al., 2002):

$$\ln L = \sum_{i=1}^n n_i \ln p_i + \sum_{i=1}^n \sum_{j=i+1}^n \frac{1}{2} n_{ij} \ln(p_i + p_j),$$

where $p_n = 1 - \sum_{i=1}^{n-1} p_i$, where n = number of the different

alleles at the particular marker, n_i = number of allele i from the founder animals (i.e., either from founder sires or unequivocally descending from unknown dams), and n_{ij} = number of half-sibs that share the same heterozygous genotype with their sire. This formula is tailored to a half-sib structure because it allows the use of genotypic information of those daughters with alleles whose paternal origin cannot be unequivocally determined.

For a single locus, the exclusion probability is defined as the probability that a putative conflict between a sire and a daughter occurs in the case of a true non-paternity and is calculated from the frequencies of the different marker alleles. Exclusion probabilities for each marker were calculated for the whole population (PE_{pop}) and each single family ($PE_{fam(i)}$). Following the S-notation given by Dodds et al. (1996), PE_{pop} was estimated for a single locus as

$$PE_{pop} = 1 - 4S_2 + 4S_3 - 3S_4 + 2S_2^2$$

where $S_t = \sum_{i=1}^n p_i^t$, where p_i = frequency of allele i , n = number of distinct alleles, and t = arbitrary non-negative integer (Dodds et al., 1996).

The family-specific exclusion probability ($PE_{fam(i)}$) was estimated following Ron et al. (1996):

$$PE_{fam(i)} = (1 - q_i)^2$$

where q_i is the sum of the frequencies of the 2 alleles of sire i .

Extending this to multiple loci, the method described by Ron et al. (1996) was followed to calculate a combined exclusion probability for the whole population (CPE_{pop}):

$$CPE_{pop} = 1 - \prod_{j=1}^m (1 - PE_{pop(j)})$$

where m = number of genotyped loci and $PE_{pop(j)}$ = exclusion probability for the population at locus j . The combined exclusion probability for each family ($CPE_{fam(i)}$) was calculated as follows:

$$CPE_{fam(i)} = 1 - \prod_{j=1}^m (1 - PE_{fam(i)j})$$

where $PE_{fam(i)j}$ = exclusion probability for family i at locus j .

The power of this study was defined as the probability of detecting a nonpaternity given a random case of nonpaternity. In a first step, the probability of nonpaternity in the case of zero to 3 loci showing a conflict was calculated. In the next step, the power was calculated as the sum of the probabilities showing a conflict at more >2 loci given a random case of nonpaternity. These probabilities can be calculated from the individual marker exclusion probabilities. One minus these probabilities of all included loci is equal to the power of this study.

Effect of Pedigree Errors on Selection Response

Following Mrode (1996), the selection response (ΔG) per generation can be defined as

$$\Delta G = i \times \sqrt{R} \times \sigma_a,$$

where i = intensity of selection, σ_a = additive genetic standard deviation, and R = reliability of the estimated breeding value; therefore, \sqrt{R} represents its accuracy.

Assuming that only progeny records contribute to the EBV of a sire, R is defined as (Mrode, 1996)

$$R = \frac{N}{(N + \lambda)}, \text{ where } \lambda = \frac{(1 - t)}{t} \quad [1]$$

where N = number of potential progenies per sire and t = intraclass correlation, which is one-quarter of the heritability in the case of a half-sib progeny group.

The impact of WSI and MSI can be illustrated by a small example. Assume a population consisting of 2 paternal half-sib groups, each of 100 daughters. Further, assume a WSI and an MSI, both of 10%. In this case, 90 daughters are assigned to each sire (MSI = 0.10), but only 81 daughters are correctly assigned to each sire (WSI = 0.10), and 9 daughters are assigned to the wrong sire, respectively. Similar to Visscher et al. (2002) it was assumed that i and σ_a were not affected

by WSI. Therefore, the impact of WSI and MSI on the selection response can be expressed as

$$\Delta G \propto \sqrt{R_e}, \text{ where } R_e = \frac{(1 - MSI) \times N}{((1 - MSI) \times N) + \lambda_e} \text{ and} \\ \lambda_e = \frac{1 - (1 - WSI) \times t}{(1 - WSI)^2 \times t}. \quad [2]$$

Subscript e denotes pedigree errors. The efficiency of a breeding plan with respect to the genetic gain with pedigree errors relative to a situation with no pedigree errors can be estimated as (Visscher et al., 2002):

$$E_G = \sqrt{\frac{R_e}{R}}, \quad [3]$$

where R is the reliability without pedigree errors.

To investigate the impact of different WSI and MSI on the response to selection, the efficiency was calculated for a number of configurations that might reflect a general dairy cattle breeding scheme based on progeny testing, including the situation that can be found in the Angeln population. Different values for heritability ($h^2 = 0.10, 0.25, \text{ and } 0.50$) and for the progeny group size ($n = 10, 50, \text{ and } 100$) were used in the calculations. The proportion of WSI was varied in 6 steps (WSI = 0.05, 0.07, 0.10, 0.15, 0.20, and 0.30), and for MSI, 4 different values were chosen (MSI = 0.10, 0.20, 0.30, and 0.40).

Additionally, as suggested by an anonymous reviewer, the different impacts of MSI and WSI on the efficiency can be derived analytically by taking the first derivative of the square of Equation (1) with respect to $(1 - MSI)$ and to $(1 - WSI)$, respectively (see Appendix).

RESULTS AND DISCUSSION

Heterozygosity and Exclusion Probability

Table 1 has the observed heterozygosities and the exclusion probabilities for the different markers. The number of alleles varied from 7 for marker BP7 to 13 for marker TGLA227. The heterozygosity of the 16 different markers ranged from 0.49 for marker INRA134 to 0.80 for markers TGLA227 and BMS2639. The exclusion probability increased with increasing heterozygosity, although this relationship was not strictly monotonic. Accordingly, the lowest exclusion probability was obtained for marker INRA134 ($PE_{pop} = 0.13$), and the highest probabilities were observed for markers TGLA227 and BMS2639 ($PE_{pop} = 0.44$). The combined exclusion probability for all 16 markers was 0.999. These results are in accordance with different studies,

Table 1. Degree of observed heterozygosity, exclusion probability (PE_{pop}), and combined exclusion probability (CPE_{pop}) in the Angeln population

| Marker | Alleles (no.) | Heterozygosity | PE_{pop} |
|-------------|---------------|----------------|------------|
| BP7 | 7 | 0.67 | 0.26 |
| RM209 | 12 | 0.63 | 0.24 |
| BMS1675 | 8 | 0.77 | 0.39 |
| BM3507 | 11 | 0.78 | 0.43 |
| CSSM028 | 11 | 0.74 | 0.34 |
| BM4513 | 9 | 0.69 | 0.29 |
| TGLA227 | 13 | 0.80 | 0.44 |
| HUJ625 | 8 | 0.76 | 0.37 |
| INRA048 | 9 | 0.76 | 0.36 |
| RM180 | 9 | 0.65 | 0.25 |
| DIK082 | 11 | 0.79 | 0.42 |
| BM6425 | 10 | 0.72 | 0.32 |
| BMC4203 | 10 | 0.73 | 0.33 |
| BM6507 | 10 | 0.65 | 0.25 |
| BMS2639 | 12 | 0.80 | 0.44 |
| INRA134 | 8 | 0.49 | 0.13 |
| CPE_{pop} | | | 0.999 |

which pointed out that a high heterozygosity is better for paternity verification, because markers with low heterozygosity are the reason for underestimating mis-identification rates (Dodds et al., 1996; Ron et al., 1996; Visscher et al., 2002).

In addition, the exclusion probabilities within single families ($PE_{fam(i)}$) were of interest (Table 2). These exclusion probabilities ranged between 0.0004 and 0.94 for the different markers and the different families. Table 2 shows that markers with a high PE_{pop} (e.g., BMS2639) did not automatically have a high exclusion

probability within each of the 5 families. Furthermore, it was shown that the $PE_{fam(i)}$ values for a given marker varied substantially within the other families (e.g., BM3507).

Adopting the classical test statistic theory to the problem of a marker-based detection of WSI, a Type I error occurs if a correct sire is erroneously declared as wrong because of genotyping mistakes, for example. A Type II error occurs if a wrong sire is not declared as one. Further, as previously mentioned, the power of the study is the probability to detect a wrong sire as being wrong. For the calculation of the power of this study, the results of Table 1 were used. The power would be the highest (>0.99) if a wrong paternity was declared when at least one locus showed a conflict. However, this would result in a relatively high Type I error rate, although measurement of the Type I error is not possible. In the case of 2 loci showing a conflict, the power would be 0.94. To find a compromise between the power and the Type I error rate, a wrong paternity was declared if a conflict was observed at ≥ 3 loci, corresponding to a power of 0.83.

Proportion of WSI

Table 3 shows the putative WSI of the Angeln population for varying numbers of conflicts. Of 805 animals, 10.8% showed conflicts at ≥ 1 loci. None of the animals showed conflicts for ≥ 8 of the 16 markers. The final WSI in the Angeln population was estimated to be 7%, bearing in mind that an unknown proportion of the WSI might be erroneously declared as WSI (i.e., reflecting a

Table 2. Exclusion probability ($PE_{fam(i)}$) and combined exclusion probability ($CPE_{fam(i)}$) within different families in the Angeln population

| | N_{gen}^1 | $PE_{fam(i)}$ | | | | |
|-------------|-------------|---------------|----------|----------|----------|----------|
| | | Family 1 | Family 2 | Family 3 | Family 4 | Family 5 |
| BP7 | 732 | 0.0004 | 0.12 | 0.0004 | 0.08 | 0.0004 |
| RM209 | 685 | 0.77 | 0.13 | 0.09 | 0.08 | 0.03 |
| BMS1675 | 604 | 0.64 | 0.55 | 0.19 | 0.24 | 0.24 |
| BM3507 | 729 | 0.25 | 0.34 | 0.64 | 0.49 | 0.92 |
| CSSM028 | 721 | 0.09 | 0.10 | 0.10 | 0.40 | 0.26 |
| BM4513 | 739 | 0.07 | 0.07 | 0.22 | 0.07 | 0.07 |
| TGLA227 | 725 | 0.31 | 0.40 | 0.31 | 0.23 | 0.94 |
| HUJ625 | 735 | 0.46 | 0.61 | 0.12 | 0.74 | 0.28 |
| INRA048 | 694 | 0.52 | 0.52 | 0.29 | 0.58 | 0.30 |
| RM180 | 734 | 0.004 | 0.07 | 0.07 | 0.14 | 0.004 |
| DIK082 | 736 | 0.34 | 0.14 | 0.23 | 0.61 | 0.34 |
| BM6425 | 735 | 0.41 | 0.07 | 0.10 | 0.29 | 0.29 |
| BMC4203 | 740 | 0.09 | 0.61 | 0.31 | 0.41 | 0.09 |
| BM6507 | 738 | 0.11 | 0.07 | 0.52 | 0.11 | 0.004 |
| BMS2639 | 743 | 0.18 | 0.14 | 0.21 | 0.46 | 0.35 |
| INRA134 | 650 | 0.16 | 0.16 | 0.05 | 0.16 | 0.05 |
| CPE_{fam} | | 0.998 | 0.996 | 0.987 | 0.999 | 0.999 |

¹ N_{gen} is the total number of genotyped daughters for the respective marker.

Table 3. Putative wrong sire information (WSI) for a varied number of conflicts in the Angeln population (Note that the final choice of the number of conflicts for declaring a WSI was ≥ 3 loci, which corresponds to a final WSI of 7% in the Angeln population.)

| Markers showing conflicts | Daughters (no.) | Putative WSI | SE ¹ |
|---------------------------|-----------------|--------------|-----------------|
| ≥ 1 | 87 | 0.108 | 0.0109 |
| ≥ 2 | 66 | 0.082 | 0.0097 |
| ≥ 3 | 56 | 0.070 | 0.0090 |
| ≥ 4 | 46 | 0.057 | 0.0082 |
| ≥ 5 | 36 | 0.045 | 0.0073 |
| ≥ 6 | 20 | 0.025 | 0.0055 |
| ≥ 7 | 5 | 0.006 | 0.0027 |
| ≥ 8 | 1 | 0.001 | 0.0011 |

¹SE = $\sqrt{\text{WSI}(1 - \text{WSI})/N}$, with n = 805 daughters.

Type I error). For the 5 different families, the estimated proportions of WSI were 4.07, 5.56, 11.11, 11.56, and 5.67%, respectively. The estimated WSI in the Angeln population is in accordance with the literature reports for other breeds (Visscher et al., 2002; Ron et al., 2003; Weller et al., 2004).

Impact of WSI and MSI on Genetic Gain

The influence of different values for WSI and MSI from 0 to 30% on reliability for the case of 100 daughters per sire and the values for the heritability of $h^2 = 0.10$ and of $h^2 = 0.25$ are presented in Figure 1. With an increase in WSI and MSI, the reliability decreased. The effect of WSI and MSI was more detrimental in the case of lower heritability. The impact of the different values of the heritability and the number of daughters per sire are presented in Table 4. These results represent the putative situation in the Angeln population, i.e., a WSI

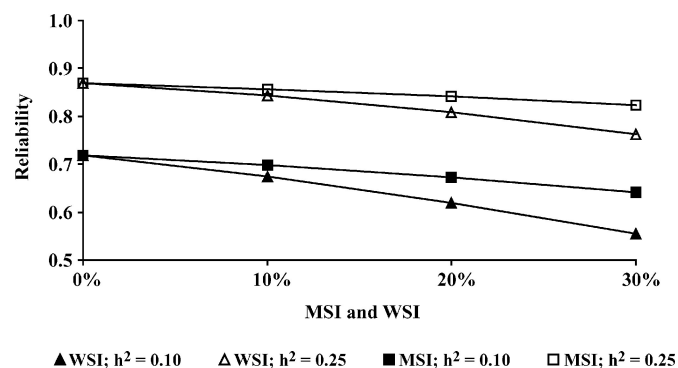


Figure 1. Reliability for the case of 100 daughters per sire, a heritability of $h^2 = 0.10$ and $h^2 = 0.25$, and different values (0 to 30%) for wrong sire information (WSI) and missing sire information (MSI).

Table 4. Reliability and efficiency of sire evaluation when there is 10% missing sire information (MSI¹) and 7% wrong sire information (WSI¹) for 3 levels of heritability and progeny size

| Heritability | Daughter group size | Reliability | Efficiency |
|--------------|---------------------|-------------|------------|
| 0.10 | 10 | 0.17 | 0.90 |
| | 50 | 0.50 | 0.94 |
| | 100 | 0.67 | 0.96 |
| 0.25 | 10 | 0.34 | 0.92 |
| | 50 | 0.72 | 0.97 |
| | 100 | 0.84 | 0.98 |
| 0.50 | 10 | 0.52 | 0.94 |
| | 50 | 0.85 | 0.98 |
| | 100 | 0.92 | 0.99 |

¹These values might reflect the situation that can be found in the Angeln population.

of 7% and a MSI of 10% (F. Reinhardt, VIT Verden, Germany; personal communication).

Other studies in the literature pointed out that WSI influenced the genetic gain downward because of the downward bias of the heritability and the lower number of progeny with correct pedigree information (Israel and Weller, 2000; Banos et al., 2001; Visscher et al., 2002). In contrast to WSI, in the study of Harder et al. (2005), MSI did not affect the estimated additive genetic variance, but the decrease in progeny size reduced the reliability of the EBV. In addition, MSI had an effect on the mean square error of the fixed effects estimate, because the estimation of the variance-covariance-matrix of the observations of the cows was incorrect for cows with MSI. Nevertheless, Harder et al. (2005) showed that it was important to have the daughters with MSI included in the estimation; otherwise, the mean square error of the fixed effects would be even greater.

The present study pointed out that WSI and MSI combined their effects on the genetic gain. The calculations showed that the impact of WSI was more harmful than that of MSI. The first derivation of Equation 1 showed, either with respect to $(1 - \text{MSI})$ or with respect to $(1 - \text{WSI})$, that the effect on the efficiency of WSI was around 1.4 times more harmful as MSI, assuming large progeny groups (see Appendix). Increasing the number of daughters per sire decreases the influence of WSI and MSI, especially in the case of a low heritability (Table 4). The results of the calculations presented in Figure 1 and Table 4 are in agreement with Van Vleck (1970a, b) and Christensen et al. (1982), who concluded that a trait with a lower heritability and WSI > 0 had higher losses in the genetic gain than traits with a higher value of heritability. Harder et al. (2005) showed similar results for MSI, which reduced the genetic gain, especially for traits with low heritabilities.

CONCLUSIONS

Microsatellite markers are suitable tools for the determination of WSI. In the German Angeln dairy cattle population, the estimated proportion of WSI was 7%. The estimation of WSI was done with the assistance of 16 microsatellite markers and accepting 3 conflicts as an indication of WSI. The power of this study, in the case of ≥ 3 loci showing a conflict, was 0.83. Additionally, it was shown that WSI and MSI had an influence on reliability and on genetic gain in the Angeln dairy cattle breed and combined their effect on genetic gain. The impact of WSI on the efficiency is around 1.4 times more harmful than the impact of MSI. If a reduction in the loss of genetic gain is desired, especially for low heritable traits caused by incorrect paternity, breeding organizations must check their recording and verification systems to decrease the proportion of wrong and missing pedigree records or to increase the number of daughters per sire.

ACKNOWLEDGMENTS

This study was supported by the Foundation Schleswig-Holsteinische Landschaft. It has benefited from the critical and very helpful comments of 2 anonymous reviewers.

REFERENCES

- Banos, G., G. R. Wiggans, and R. L. Powell. 2001. Impact of paternity errors in cow identification on genetic evaluations and international comparisons. *J. Dairy Sci.* 84:2523–2529.
- Bennewitz, J., N. Reinsch, and E. Kalm. 2002. GENCHECK: A program for consistency checking and derivation of genotypes at codominant and dominant loci. *J. Anim. Breed. Genet.* 119:350–360.
- Brka, M., N. Reinsch, and E. Kalm. 2002. Is there linkage between supernumerary teats in cattle and BTA3 markers? *Arch. Tierz.* 45:429–432.
- Christensen, L. G., P. Madsen, and J. Petersen. 1982. The influence of incorrect sire-identification on the estimates of genetic parameters and breeding values. *Proc. 2nd World Congr. Genet. Appl. Livest. Prod., Madrid, Spain VII:200–208.*
- Dodds, K. G., M. L. Tate, J. C. McEwan, and A. M. Crawford. 1996. Exclusion probabilities for pedigree testing farm animals. *Theor. Appl. Genet.* 92:966–975.
- EAAP. 2004. Online. <http://www.tiho-hannover.de/einricht/zucht/eaap/descript/8.htm>
- Harder, B., J. Bennewitz, N. Reinsch, M. Mayer, and E. Kalm. 2005. Effect of missing sire information on genetic evaluation. *Arch. Tierz.* 48:219–232.
- Israel, C., and J. I. Weller. 2000. Effect of misidentification on genetic gain and estimation of breeding value in dairy cattle populations. *J. Dairy Sci.* 83:181–187.
- Mrode, R. A. 1996. *Linear Models for the Prediction of Animal Breeding Values.* CAB International, Wallingford, Oxon, UK.
- Myakishev, M. V., G. I. Kapanadze, G. O. Shaikhayev, G. P. Georgiev, and D. R. Beritashvili. 1995. Extraction of DNA from the Whole Blood by Silica Gel. *Inst. Gene Biology, Moscow, Russia.*
- Reinsch, N. 1999. A multiple-species, multiple-project database for genotypes at codominant loci. *J. Anim. Breed. Genet.* 116:425–435.
- Ron, M., Y. Blanc, M. Band, E. Ezra, and J. I. Weller. 1996. Misidentification rate in the Israeli dairy cattle population and its implications for genetic improvement. *J. Dairy Sci.* 79:676–681.
- Ron, M., R. Domochofsky, M. Golik, E. Seroussi, E. Ezra, C. Shturman, and J. I. Weller. 2003. Analysis of vaginal swabs for paternity testing and marker-assisted selection in cattle. *J. Dairy Sci.* 86:1818–1820.
- Savaş, T., N. Reinsch, and E. Kalm. 1998. Auswirkungen der Rassenzusammensetzung auf Merkmale der Tagesmilchmenge und auf die Zellzahl beim Angler Rind. *Arch. Tierz.* 41:201–209.
- Van Vleck, L. D. 1970a. Misidentification in estimating the paternal sib correlation. *J. Dairy Sci.* 53:1469–1474.
- Van Vleck, L. D. 1970b. Misidentification and sire evaluation. *J. Dairy Sci.* 53:1697–1702.
- Visscher, P. M., J. A. Woolliams, D. Smith, and J. L. Williams. 2002. Estimation of pedigree errors in the UK dairy population using microsatellite markers and the impact on selection. *J. Dairy Sci.* 85:2368–2375.
- Weller, J. I., E. Feldmesser, M. Golik, I. Tager-Cohen, R. Domochofsky, O. Alus, E. Ezra, and M. Ron. 2004. Factors affecting incorrect paternity assignment in the Israeli Holstein population. *J. Dairy Sci.* 87:2627–2640.
- Weller, J. I., Y. Kashi, and M. Soller. 1990. Power of daughter and granddaughter designs for determining linkage between marker loci and quantitative trait loci in dairy cattle. *J. Dairy Sci.* 73:2525–2537.

APPENDIX

As shown in Equation 3 of the main text, efficiency (E_G) is defined as $E_G = \sqrt{\frac{R_e}{R}}$, where R_e = reliability with pedigree errors (Equation 2 in the main text) and R = reliability without pedigree errors (Equation 1 in the main text).

Using Equations 1, 2, and 3 and defining $x = (1 - \text{MSI})$ and $y = (1 - \text{WSI})$, the square of the efficiency is given by

$$E_G^2 = \frac{xy^2[(N - 1)t + 1]}{yt(xyN - 1) + 1} \quad [\text{A1}]$$

where N = number of potential progenies per sire and t = intraclass correlation.

Taking the first derivative of Equation A1 with respect to x yields with some algebraic operations:

$$\frac{\delta E_G^2}{\delta x} = \frac{E_G^2}{x} - \frac{y^2 N t \times E_G^2}{yt(xyN - 1) + 1} \quad [\text{A2}]$$

Setting $x = y = E_G = 1$ (i.e., no pedigree errors) results in

$$\frac{\delta E_G^2}{\delta x} = 1 - \frac{Nt}{t(N - 1) + 1} = 1 - R. \quad [\text{A3}]$$

Taking the first derivative of Equation A1 with respect to y yields, after some algebraic operations,

$$\frac{\delta E_G^2}{\delta y} = \frac{2E_G^2}{y} - \frac{E_G^2 \times [t(2xyN - 1)]}{yt(xyN - 1) + 1}. \quad [A4]$$

Setting $x = y = E_G = 1$ (i.e., no pedigree errors) results in

$$\frac{\delta E_G^2}{\delta y} = 2 - \frac{t(2N - 1)}{t(N - 1) + 1} = 2(1 - R) - \frac{t}{t(N - 1) + 1}. \quad [A5]$$

Equations A3 and A5 show that the increase of MSI and WSI (i.e., a decrease of x and y) results in a loss of E_G^2 , that is equal for $(1 - R)$ and $(2(1 - R) - \frac{t}{t(N-1) + 1})$, respectively. Hence, for large N , the impact of WSI on E_G is around 1.4 (i.e., $\sqrt{2}$) times as harmful as MSI.