

¹Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala, Sweden

²Department of Animal Science, Isfahan University of Technology, Isfahan, Iran

MOHAMMAD ALI NILFOROOSHAN¹, AMIR KHAZAELI² and MOHAMMAD ALI EDRISS²

Effects of missing pedigree information on dairy cattle genetic evaluations (short communication)

Abstract

Estimating genetic merit of livestock closest to their true genetic merit is a preliminary goal in animal breeding. The accuracy of genetic evaluations depends on the recording system and the method of evaluation. Whereas applying more complicated models may improve the accuracy of evaluations inconsiderably, improving data quality is more effective. The data were on pedigree and milk performances (milk yield, fat yield and fat percentage) of 9834 dairy cows in Isfahan, Iran, with both known parents. Genetic parameters were estimated by derivative-free restricted maximum likelihood method, applying an animal model (full relationships), sire model (dam missed), dam model (sire missed) and a half-missed model (half sire / half dam). All the models were compared to animal model. Sire model had the smallest size of pedigree structure, while dam model had an inexistence of between herd relationships. The results showed underestimating additive genetic variance by sire and half-missed models and overestimating it by dam model. An important finding of this study was that there is an unfavorable interaction between missing sire and dam information that caused the lowest goodness of fit for half-missed model. Also, usually, sire missing makes more important problems to the pedigree structure and genetic evaluations than dam missing. The research revealed that, even using an animal model, there are some delicacies in introducing the relationship matrix for sex-limited traits, which requires special attention to the pedigree of sires.

Key words: animal model, sire model, pedigree missing, sex-limited

Zusammenfassung

Titel der Arbeit: **Wirkung fehlender Abstammungsinformationen auf Zuchtwerte bei Milchrindern**

Die Schätzung der Zuchtwerte möglich sicher am tatsächlichen genetischen Potential der Tiere ist ein wesentliches Ziel in der Tierzucht. Die Genauigkeit dieser Schätzung hängt wesentlich vom Datenerfassungssystem und der Schätzmethode ab. Dabei erweist sich die qualitative Verbesserung der verfügbaren Daten als effektiver als die Wahl komplizierter Modelle. In vorliegender Untersuchung wurden Abstammungsdaten und Milchleistungen (Milchertrag, Fettertrag und Fettprozent) von 9834 Milchkühen der Isfahan Provinz im Iran mit beiden bekannten Eltern erfasst. Die genetischen Parameter wurden mit Hilfe der eingeschränkten Maximum-Likelihood-Schätzung ermittelt, zur Anwendung kamen das Tiermodell mit vollständiger Dateninformation, das Vatertiermodell (ohne Mütterinformation), das Müttermodell (ohne Vaterinformation) und ein Modell mit je der halben Elterninformation. Alle Modelle wurden mit dem Tiermodell verglichen. Das Vatertiermodell hatte die geringste Größe der Abstammungsstruktur während beim Müttermodell keine Innerherdenbeziehungen vorhanden waren. Die Ergebnisse zeigen eine Unterschätzung der additiven genetischen Varianz beim Vatertier- und den Halbinformationsmodellen und eine Überschätzung beim Müttermodell. Ein wichtiges Ergebnis dieser Studie war eine Wechselwirkung zwischen fehlender Väter- oder Mütterinformation welche die geringste Anpassungsgüte der Halbinformationsmodelle begründet. Für eine Zuchtwertschätzung erwiesen sich fehlende Vaterinformationen als problematischer als fehlende Mütterinformationen. Die Untersuchung zeigte, dass sogar bei Nutzung des Tiermodells besondere Aufmerksamkeit der Bullenabstammung bei Betrachtung der Verhältnismatrix für Geschlechter begrenzte Merkmale zu widmen ist.

Schlüsselwörter: Tiermodell, Vatertiermodell, Stammbaum, fehlende Zuchtinformation, Geschlechtsbegrenzung

Introduction

Proving dairy sires on a national basis was initiated in the United States in 1933 by the United States Department of Agriculture (USDA); proofs were from daughter-dam comparisons (WILCOX et al., 1992). Because daughter-dam regression method tends to be biased by maternal effects, it has not been widely used. Nowadays, using REML procedures, evaluations are free of this bias (NDLOVU, 1993). In 1962, the method changed to daughter-herdmate comparisons which was further improved in 1965 by taking some additional parameters into account. The MCC (Modified Contemporary Comparison) method was implemented in October 1974 in attempting to take into consideration some items including genetic differences in herdmates, pedigree information, daughter distribution over herds, length of lactation, number of herdmates, and the number and average repeatability of herdmate sires (WILCOX et al., 1992). By the invention of Henderson's methodology of best linear unbiased prediction (BLUP), sire, sire-maternal grandsire, and animal models were invented and investigated. The techniques are well accepted and BLUP nowadays is the preferred method of evaluation (HERRENDÖRFER et al., 1999; WILCOX et al., 1992). Initially, evaluations were taking from a best linear unbiased prediction (BLUP) sire model. Then, it was improved by adding maternal grandsires to the pedigree structure (sire-MGS model) in order to improve the accuracy by considering more genetic relationships, direct effect as sire and maternal effect as sire of dam, partial accounting of merit of mates, and differences in maternal ability of dams. In February and August 1989, animal model proofs became available for type and production, respectively (BURNSIDE et al., 1989). Animal and sire models are very similar. However, animal model considers all relatives, evaluates sires and dams simultaneously and therefore genetic merits are adjusted for any non-random mating and as a consequence it is more accurate (BURNSIDE et al., 1989). Thanks to HENDERSON (1975) for the discovery of the algorithm for finding the inverse of the relationship matrix; nowadays genetic evaluations have been based on restricted maximum likelihood (REML) methods. New algorithms such as derivative-free restricted maximum likelihood (DFREML) and the relevant computer programs have been developed (GRASER et al., 1987; MEYER, 1998) with the aim of more precise and unbiased evaluations. Improvement of the computational power of computers has made it possible to perform more complicated animal models, and as well as single-trait animal models, multi-trait and random regression animal models are in use in various researches (ATIL et al., 2005; KREJČOVÁ et al., 2007a; KREJČOVÁ et al., 2007b).

CANTET et al. (2000) using a simulated data, studied the different situations of: (a) complete pedigree; (b) 50% of phenotypes with sire missing; (c) 50% of phenotypes with dam missing (d) and 50% of phenotypes with sire and dam missing (the paternities of 12.5%, the maternities of 12.5%, and the paternities and maternities of 12.5% of the records were lost at random). Choice (a) produced unbiased heritabilities very close to the true heritabilities. Heritability estimates were more biased when 50% of sires were missing (b) than when 50% of dams were missing (c) and choice (d) was the most biased. Choice (a) which is the ideal case was termed "ignorable selection" and the other choices were termed "non-ignorable selection" by IM et al. (1989). They concluded that if all data employ in making selection decisions, the selection process may be ignored and estimation may proceed as if selection has never been occurred.

REML estimates are not biased by “ignorable selection” (IM et al., 1989; CANTET et al., 2000).

There are two kinds of pedigree errors affecting the results of genetic evaluations, including wrong and missing pedigree information (HARDER et al., 2005). The impact of pedigree errors on reducing genetic gain is large and the effect of wrong pedigree is 1.4 times more harmful than the effect of missing pedigree (SANDERS et al., 2006). Missing parents leads to serious underestimation of inbreeding and therefore making necessary decisions against inbreeding may be delayed (LUTAAYA et al., 1999; HARDER et al., 2005). LUTAAYA et al. (1999) simulated various percentages of dam missing information and reported considerable underestimation of inbreeding in the population. In order to decrease the proportion of unknown paternity and increase the average number of daughters per sire, breeding organizations should recheck their recording and verification systems (SANDERS et al., 2006).

Although in developed countries missing pedigree data (especially sires) is a problem for beef cattle populations, in developing countries, missing record and pedigree data has remained a serious problem for dairy cattle populations. However, even developed countries are not free of these errors (HARDER et al., 2005; SANDERS et al., 2006). While models of evaluation become more complicated just to increase the accuracy a little bit, improving data quality can do all that complicated models cannot do. Applying complicated models also has some problems; they are more data specific and computational demanded and may not be well fitted to new data. This study will show how collecting good data by dairy farmers, Animal Breeding Centers or recording agencies is important in improving the accuracy of genetic evaluations. The objectives of this study were: 1) Comparison between animal model and sire model 2) Comprehensive theoretical and experimental investigations on the effects of missing pedigree information; the ways that missing sire or dam each will damage pedigree structure; and special cases for sex-limited traits 3) Studying the effect of missing pedigree on the accuracy of estimating variance components and breeding values 4) How missing sire or dam data acts on estimating genetic parameters for traits with different heritabilities.

Materials and methods

Data

The data were on pedigree and mature equivalent standardized (ME-305d-2X) first lactation records of Holstein cows, collected by Animal Breeding Center of the Ministry of Agriculture in Isfahan, Iran, from 35 dairy herds from 1994 to 2002. The range of age at first calving was considered to be between 21 to 39 months and the remaining part of the data was excluded. Since the database itself should not be under the effect of missing pedigree information, animals with records were enforced to have both sire and dam identification numbers. Therefore, animals with unknown sire or dam were not permitted to be included in the data set on which we wanted to run animal model. The final data were consisted of pedigree, Herd-Year-Season of calving, milk yield, fat yield and fat percentage records of 9834 dairy cows.

Analysis

Intentionally, dam and sire pedigree information were excluded to study sire model (dam missed) and dam model (sire missed), respectively. Although dam model has

never been a conventional model for dairy cattle evaluation, it was considered just to simulate sire missing situation. In addition, it is a more complex version of daughter-dam regression and daughter-herdmate methods, equipped to a dam-based relationship matrix. To study the effect of both missing sire and dam pedigree information, another model called half-missed model (half sire / half dam) was created, for which the phenotypes were assigned randomly into two groups. For one group, paternity information and for the other maternity information were lost. For simplicity, the situation of full relationships (no missed data) was called animal model. However, animal model is also in use for data bases with missing values, and half-missed model itself is a kind of animal model. National genetic evaluations involve animals with incomplete pedigrees (LUTAAAYA et al., 1999). The following model was employed to obtain variance components and genetic parameters using Animal Model (AM), Sire Model (SM), Dam Model (DM) and Half-missed Model (HM).

$$Y = Xb + Za + e$$

Where: Y , b , a , and e are the vectors of observations, fixed contemporary groups (HYS), random direct additive genetic, and random residual effects, respectively. X and Z are the incidence matrices relating records to fixed and direct additive genetic effects, respectively. Due to the lower number of records for fat traits (401 less records), the number of HYS levels was lower (464 vs. 473) relative to milk yield.

All models were run under DFREML 3.0 β package (MEYER, 1998) using derivative-free restricted maximum likelihood method by Simplex way, applying univariate procedure (DFUNI). Convergence criterion was set on 10^{-8} , and DFREML (MEYER, 1998) automatically pruned all single-linked parents. Finally, estimated breeding values of animals for different traits applying various models were compared using SAS statistical software (SAS Inst., 1997).

Results and discussion

Pedigree missing, different situations and consequences

AM considers all information available on all relatives to increase the accuracy of evaluations. All ancestors and descendants are used in the evaluation of both sires and cows, weighted by how closely related they are. It appropriately adjusts for merit of mates to rank sires. AM evaluates all animals simultaneously by adjusting equations for management and environmental factors and solving them to evaluate genetic merit of animals (WILCOX et al., 1992). SM and DM use just certain classes of relatives, because relationships are defined through sire and dam lines, respectively, and there would be no correction for merit of mates. In AM, full-sib and half-sib relationships are known. For example, it is clear which animals are full-sibs or paternal/maternal half-sibs with each other. This is also true for full-cousins, half-cousins and, etc., whereas in SM and DM there is no difference between full-sibs and half-sibs (i.e. all of them are considered as paternal and maternal half-sibs by SM and DM, respectively). For SM and DM, there would be a decline in the average number of grand progeny per grandparent, because applying SM, sires will lose their daughters' progeny, and applying DM, dams will lose their sons' progeny. By DM, there is an inexistence of between herd genetic relationships, because in dairy cattle industry, progeny of each dam are usually reared and milked in the same herd and the rate of heifer exchange is negligible, while most sires establish genetic connections between

herds because sires have progeny from various dams in various herds and relationship connections between herds pass through sires (see Appendix). In AM, in addition to well-established genetic relationships between herds, within herd relationships become strong by dams. Missing pedigree data may reduce the accuracy of correction genetic merits for fixed effects because it becomes less clear how animals in different fixed groups (e.g., herds or contemporary groups) are genetically similar or different. The more herds or contemporary groups are genetically similar, the more their differences would be environmental.

In each model of evaluation, at least one fixed effect is included, which for dairy cattle evaluations it is usually Herd-Year-Season for contemporary grouping. Well defined genetic ties within and between contemporary groups are of interest because comparison of animals becomes possible. In AM, relationship matrix can make a better connection with the matrix related to fixed parts of the model (mainly, contemporary groups) because better defined genetic relationships are between and within various fixed groups. Missing pedigree information would lead to weakening both within and between contemporary group genetic relationships. It seems that losing dam information decreases within herd contemporary group genetic ties, and losing sire information decreases between herd contemporary group genetic ties. Also, it is likely that losing pedigree information leads to weakening/gaps between herds and generations, especially by missing sires and dams, respectively (see appendix). Since sires mate to various cows for several generations, missing sire information cuts more relationships than missing dams, which calve a few progeny during their productive life. By missing a sire, in addition to replacing full-sib relationships with maternal half-sibs, animals will lose their paternal half sibs, which are more than their maternal half-sibs. By increasing the proportion of missing sire information, many half-sib cows are treated as unrelated. Therefore, the (co)variance matrix of observations is estimated incorrectly for data sets with missing sire information (HARDER et al., 2005). It seems that sire missing would make pedigree narrower in width, whereas dam missing would make pedigree shallower in depth. In extreme conditions, the number of loops will increase in the pedigree. In AM, because animals make various relationships with each other through sires and dams, pedigree loop(s) become wider, deeper and lower in the potential number (usually to one), containing more animals, leading to more accurate and reliable evaluations.

The accuracy of the genetic evaluation of an animal depends on the number of its progeny in different herds, the number of its full/half-sibs and their progeny, whether the animal itself has a missing parent, how its parents had been involved in the pedigree structure, the number of available records on the animal itself and its relatives (especially, the closer ones) as well as the heritability of the trait.

Pedigree study

A summary of the pedigree structure obtained by different models is represented in Table 1. Considering the size of the pedigree structure (no. of animals in the model), AM had the highest and SM had the lowest defined population due to the lower number of sires mated to more dams. There were 29 single-linked sires and 3,874 single-linked dams in the pedigree, which were pruned automatically by DFREML (MEYER, 1998) to improve the model's efficiency. The number of animals with pruned dam was substantially higher, due to the high number of cows with one (female) progeny, and many cows with more than one female progeny, but only one

milking and recorded. Animals with only one progeny and no records were considered to be pruned because they do not contribute to any information or relationship. The number of single-linked parents in HM was a balance between the number of missed single-linked parents and the number of multi-linked parents, which became single-linked by HM.

Since the traits were sex-limited, the pedigree could not be defined for animals without records (sires), as a result, sires compelled to be genetically independent. Thus, only daughter-sires and paternal half-sib sisters were identified by SM (see Appendix). There was no sire/grandsire and paternal grand dam in DM, and no dam, grandparent/great-grandparent in SM. The decrease in the number of great-grandparents from AM to DM was only due to missing maternal grandsires. The surprising point about SM was that the pedigree depth did not reach to any grandparent. The reason was that for sex-limited traits by failure in introducing pedigree information for sires, even using AM, no paternal grandparent can be identified in the pedigree, and by inconsideration of maternal relationships (SM), there could not be any maternal grandparent as well as any paternal grandparent. The number of great-grandparents considerably declined (5.7 times) from AM to HM. Considering the low reduction (1.4 times) in the number of grandsires, this decline was mainly due to the high reduction (5 times) in the number of grand dams (Tab. 1). There were no quarter-sibs in the pedigree structure for SM, because no son-parent relationship could be identified for sex limited traits, and there were no quarter-sibs contributed by sons as future sires. Thus, there could not be any quarter-sib without any dam (SM). In this situation, SM uses records to evaluate the animal itself, its sire and its paternal sisters (from all kinds of record pairs, SM contained only paternal half-sib record pairs (Tab. 1)). Thus, using SM would not be precise, especially for sex-limited traits. Although AM and HM differ in half of the pedigree information, the number of half-sib and quarter-sib record pairs was 3.9 and 18 times lower in HM, respectively, which shows how sires and dams can complete the relationship net with each other.

Although HM had 84 quarter-sib record pairs that SM did not have, the number of half-sib record pairs were 208,496 for HM relative to 806,520 for SM. This comparison shows that the more half paternity information in SM is worthier the half maternity information in HM. Comparing SM and DM, SM had no identified grandparent and quarter-sib (1,119 and 1,264 record pairs, respectively for DM). However, in addition to a better genetic relationship between management groups, SM had 125.7 times more half-sib record pairs relative to DM. This again confirms the importance of paternity information relative to the maternity one.

For AM, there would be no limitation in introducing relationships between animals. However, for sex-limited traits this freedom may become restricted to the sire sides for several reasons. While the main concern is on the availability of pedigree for animals with records, providing pedigree information for animals without records (both sires and dams) may be neglected. Although it does not make any important problem for sex-unlimited traits, for sex-limited traits it leads to genetic unrelatedness between animals without records, especially sires, which are of the most importance. Generally, dairy farmers are responsible for providing and submitting records and pedigree information of their cows. Thus, it seems reasonable that animal breeding organizations pay more attention to provide pedigree information of sires in their

genetic evaluations. Also, there are some differences between different computer packages as well as different procedures within packages. For example, in the case of this study, in the single-trait procedure of DFREML package (DFUNI), a pedigree-record input file has to be provided. Thus, the pedigree can be defined for animals with records. However, applying multi-trait or repeated records procedures (DXMUX and DXMRR, respectively), two input files including one pedigree-record and a pedigree file have to be provided. The aim of providing a separate pedigree file is that the pedigree in the pedigree-record file is mixed for various traits to relate records in different traits to animals and various traits may differ in the availability of some parts of the pedigree. Thus, the pedigree input file is to have a combined pedigree for various traits and put them in a relationship matrix. This can provide a good opportunity for sex-limited traits to compensate their pedigree incompleteness through a pedigree contributed by a sex-unlimited trait in a multi-trait analysis or with another sex-limited trait, providing pedigree for animals without records (at least for sires) in the pedigree input file. Therefore, the possibility of providing pedigree information for animals without record in the package/procedure is critical for performing genetic evaluations for sex-limited traits.

Table 1

Summary of the pedigree structure for different models (Zusammenfassung der Zuchtstruktur für unterschiedliche Modelle)

Data and pedigree structure	AM	SM	DM	HM
no. of animals in total	15566	10174	15226	13227
... with own record	9834	9834	9834	9834
... with unknown/pruned sire	29 ^P	29 ^P	9834	4873+40 ^P
... with unknown/pruned dam	3874 ^P	9834	3874 ^P	4961+2579 ^P
... in the model	11663	10145	11352	10608
... with great-grand parents	189	0	100	33
no. of sires	311	311	0	275
... paternal half-sib record pairs	806520	806520	0	206816
... which are also paternal grandsires	0	0	0	0
... which are also maternal grandsires	184	0	0	134
... quarter-sib record pairs	78894	0	0	4360
no. of dams	3460	0	3460	1573
... with own record	1942	0	1942	1074
... dam-offspring record pairs	2441	0	2441	1200
... maternal half-sib record pairs	6414	0	6414	1680
... which are also paternal grand dams	0	0	0	0
... which are also maternal grand dams	935	0	935	188
... with own record	170	0	170	47
... grand dam-offspring record pairs	190	0	190	48
... quarter-sib record pairs	1264	0	1264	84

P: Animals with pruned parent.

Variance components

Heritabilities and the variance components estimation are presented in Table 2. For all of the studied traits, additive genetic variances (V_A) and consequently heritabilities (h^2) were underestimated by SM and HM and overestimated by DM. This may be due to the higher selection intensity of males than females, and genetic independency between herds for DM. CANTET et al. (2000) reported biased downward heritability estimates by loss of pedigree information. DONG et al. (1988) studied the degree of completeness of relationships and reported lower heritabilities by REML if relationships are from sires only, compared to those from more complete pedigrees. They also found that full relationships from ancestors of about two generations result in slightly higher heritabilities than when relationships are from only one generation.

Considering $|h^2_{AM}-h^2_{SM}|/h^2_{AM}$, $|h^2_{AM}-h^2_{DM}|/h^2_{AM}$ and $|h^2_{AM}-h^2_{HM}|/h^2_{AM}$ for milk yield (0.414, 0.516 and 0.332, respectively), fat yield (0.268, 0.327 and 0.254, respectively) and %fat (0.261, 0.122 and 0.007, respectively), the heritability estimates by HM were the closest to AM. This may be due to the closest ratio of males/females relative to AM, a slight correction for merit of mates, and considering genetic variation between sires (relative to DM). Due to the inconsideration of genetic variation between sires, heritability estimates by DM were more different to heritability estimates by AM. The ratios were the highest for fat percentage and the lowest for milk yield. This helps to make this conclusion that the damage from lost pedigree is lower for high heritable traits. Pedigree missing makes more problems for traits with lower heritability and in the case of small progeny size per sire (SANDERS et al., 2006).

In terms of precision, the lowest Log likelihood and the highest heritability standard error were for HM, showing the lowest precision for it. This can be a clue for an unfavorable sire by dam missing interaction on the accuracy of evaluations. As it was expected, SM had a better goodness of fit (from both Log likelihood and SE of h^2 points of view) relative to DM.

Using AM, heritabilities were in the range of other studies. EDRISS et al. (2006), using the same database including additional data from animals with known/unknown parent(s), reported 0.229 and 0.242 heritabilities for milk and fat yields, respectively. Also, NILFOROOSHAN and EDRISS (2007), using (305d-2X) records of animals with known sire and known/unknown dam (unknown maternity for %4 of the data, not reported previously) and age at calving as a covariable in the model, estimated 0.24, 0.27 and 0.41 heritabilities for milk yield, fat yield and %fat, respectively. KHATTAB et al. (2005), using two animal models, estimated the heritability of 305d milk yield of Friesians in Egypt between 0.22 and 0.23.

Table 2

Estimates of variance components, applying different models for the studied traits (Varianzkomponentenschätzung der untersuchten Merkmale bei unterschiedlichen Modellen)

Trait	Model	V_A	V_E	V_P	$h^2 \pm S.E.$	Log L.
Milk Yield (kg)	AM	443905.4	1374884.7	1818790.1	0.244 ± 0.033	-72511
	SM	255861.1	1530002.6	1785863.7	0.143 ± 0.037	-72545
	DM	657635.3	1120158.1	1777793.4	0.370 ± 0.040	-72554
	HM	289947.9	1491441.4	1781389.3	0.163 ± 0.037	-72574
Fat Yield (kg)	AM	329.36	882.92	1212.28	0.272 ± 0.031	-36597
	SM	237.64	958.27	1195.91	0.199 ± 0.032	-36636
	DM	427.60	757.56	1185.16	0.361 ± 0.038	-36670
	HM	242.22	949.16	1191.38	0.203 ± 0.040	-36686
%Fat	AM	0.0671	0.0908	0.1579	0.425 ± 0.032	3657
	SM	0.0484	0.1057	0.1541	0.314 ± 0.041	3575
	DM	0.0732	0.0802	0.1534	0.477 ± 0.038	3453
	HM	0.0654	0.0896	0.1550	0.422 ± 0.045	3489

V_A , V_E and V_P are direct additive genetic, error, and phenotypic variances, respectively. Log L. is the Logarithm of the restricted maximum Likelihood function

Genetic evaluations

Table 3 shows the ranges in which breeding values were predicted by different models. HM and SM estimated breeding values in a narrower range, whereas DM estimated them in a wider range relative to AM. This can be due to the underestimation and overestimation of heritabilities by SM and DM, respectively (Tab. 2). Using a SM, HARDER et al. (2005) also overestimated DYD variance as a result of sire missing information. Whereas heritabilities were higher for HM than SM, the ranges of the estimated breeding values were lower in HM due to lower V_P estimates by HM.

The models concerned estimated generalized least squares for contemporary groups (HYS) in a relatively close range. Missing pedigree information directly affects **Z** matrix, which will form the additive relationship matrix, and it has nothing to do with **X** matrix by which animals assign to fixed groups. However, slight differences in fixed effect solutions are because of changes in the shape of likelihood surface due to lost pedigree.

Correlation coefficients between breeding values by AM and the other models with missed data are represented in Table 4. SM had the highest and DM had the lowest correlations with AM. The average correlation between AM and SM was 0.984 relative to 0.817 for HM, which shows that although sires ranked similarly by SM as they ranked by AM, as a result of losing half of daughters, sires re-ranked in a different scale by HM. As a result of missing sire information, decreasing the number of daughters per sire leads to a reduction in response to selection, especially for low heritable traits (HARDER et al., 2005). BURNSIDE et al. (1989) reported that sires rank similarly for type traits between SM and AM due to the high estimated correlations (0.97) between sire ratings by SM and AM that were very close to +1, which would be the correlations if sires rank exactly the same. BOETTCHER et al. (1999) estimated lower correlations (0.87) between estimated breeding values from SM and AM for survival. Also, breeding value and ranking correlations are in use between different animal models (e.g., multi-trait and random regression models) to see how they evaluate and rank animals differently (KREJČOVÁ et al., 2007_(a)).

Correlations were the highest for fat percentage and the lowest for milk yield, which showed that the problem caused by missing pedigree data may be lower for higher heritable traits. This conclusion is in agreement with the results shown in Table 2. HARDER et al. (2005) using a simulated data and SM with different proportions of sire missing, estimated lower rank correlations and more loss of response to selection by more sire missing and lower heritability.

Table 3

Ranges of estimated breeding values (EBV) and estimated Herd-Year-Season generalized least squares, applying various models (Rangfolge der geschätzten Zuchtwerte und Herde-Jahreffeekte bei unterschiedlichen Modellen)

Trait	Model	N	EBV Range (std.)	HYS GLS Range (std.)
Milk Yield (kg)	AM	11663	3091.6 (348.9)	6753.4 (955.9)
	SM	10145	2213.8 (247.3)	6801.2 (953.5)
	DM	11352	3929.5 (478.7)	6678.1 (951.9)
	HM	10608	1962.7 (235.1)	6672.8 (955.5)
Fat Yield (kg)	AM	11192	95.11 (10.07)	174.31 (30.64)
	SM	9726	76.67 (8.21)	172.25 (30.60)
	DM	10885	98.77 (12.11)	179.94 (30.62)
	HM	10171	64.90 (7.39)	178.27 (30.61)
%Fat	AM	11209	1.541 (0.170)	2.695 (0.358)
	SM	9740	1.150 (0.140)	2.700 (0.356)
	DM	10902	1.770 (0.180)	2.655 (0.352)
	HM	10187	1.545 (0.165)	2.709 (0.354)

Mean differences between estimated breeding values by AM and the other models with missed pedigree data were compared by performing T-tests using SAS software (SAS Inst., 1997) to study how missing data can make evaluations different ($H_0: \mu = 0$). The results (Tab. 4) showed that, except for the deviations for HM in milk and fat yields for sires and for DM in milk yield, missed pedigree data in the studied situations (SM, DM and HM) made evaluations significantly different from the full relationships situation (AM). Although evaluations made by HM in milk and fat yields for sires, and

DM in milk yield were not significantly different from AM evaluations, the related correlations were of the lowest, which shows that the direction of changes in evaluations were not the same with AM.

Table 4

Correlations and mean differences between estimated breeding values applying various models (Korrelationen und Mittelwertunterschiede zwischen den geschätzten Zuchtwerten bei unterschiedlichen Modellen)

Trait	Gender	Variables	N	Correlation	Deviation	T-dev ^α
Milk Yield (kg)	Sires	AM, SM	311	0.979	24.6	3.63***
		AM, DM	0	—	—	—
		AM, HM	275	0.775	12.3	0.75 n.s.
	Cows	AM, SM	9834	0.935	23.1	14.66***
		AM, DM	11352	0.867	2.2	0.93 n.s.
		AM, HM	10333	0.883	10.3	5.66***
Fat Yield (kg)	Sires	AM, SM	307	0.987	0.34	2.57**
		AM, DM	0	—	—	—
		AM, HM	269	0.827	-0.39	-0.92 n.s.
	Cows	AM, SM	9419	0.951	0.30	8.05***
		AM, DM	10885	0.872	-0.40	-7.02***
		AM, HM	9902	0.895	-0.38	-7.59***
%Fat	Sires	AM, SM	307	0.987	-0.011	-5.93***
		AM, DM	0	—	—	—
		AM, HM	269	0.850	-0.017	-2.77**
	Cows	AM, SM	9433	0.960	-0.010	-16.63***
		AM, DM	10902	0.919	-0.009	-13.01***
		AM, HM	9918	0.936	-0.013	-21.19***

α H0: $\mu = 0$; *** $p < 0.001$; n.s. $p > 0.05$

Conclusion

Missing pedigree data would decrease or even detach some relationships between animals and reduces the power of animal models in genetic evaluation. Missing sire pedigree information is more harmful than missing dam pedigree information, both for the evaluation of sire itself by reducing the number of its daughters, and also for its daughters by making them unrelated or replacing full-sib with half-sib relationships. Missing pedigree information also reduces the ability of animal models to correct for the merit of mates. There are evidences regarding an unfavorable interaction of sire by dam missing. Except for heritability (and not its standard error), HM was below than an intermediate between SM and DM in all of the studied aspects. Performing genetic evaluations for dairy cattle, it is important providing pedigree information for sires as animals without records (for most of the economical traits) to make possible relationships between them. Consequently, sires can benefit from each other evaluations. The effect of pedigree missing information depends on the rate of sire missing, the distribution of the pedigree between herds and whether AI in use, the rate of dam missing, the depth of the pedigree and whether the trait concerned is sex-limited.

References

- ATIL, H.; KHATTAB, A.S.; BADAWY, L.:
Genetic parameter of birth and weaning weights for Friesian calves by using an animal model Arch. Tierz., Dummerstorf **48** (2005), 261-269
- BOETTCHER, P.J.; JAIRATH, L.K.; DEKKERS, J.C.M.:
Comparison of methods for genetic evaluation of sires for survival of their daughters in the first three lactations. J. Dairy Sci. **82** (1999), 1034-1044

- BURNSIDE, T.; SULLIVAN, B.; SMITH, G.:
The animal model – making cow family genetics work to the breeder's advantage. *Holstein J.* **52** (1989), 25
- CANTET, R.J.C.; BIRCHMEIER, A.N.; SANTOS-CRISTAL, M.G.; DE AVILA, V.S.:
Comparison of restricted maximum likelihood and Method R for estimating heritability and predicting breeding value under selection. *J. Anim. Sci.* **78** (2000), 2554-2560
- DONG, M.C.; VAN VLECK, L.D.; WIGGANS, G.R.:
Effect of relationships on estimation of variance components with an animal model and restricted maximum likelihood. *J. Dairy Sci.* **71** (1988), 3047-3052
- EDRISS, M.A.; NILFOROOSHAN, M.A.; SADEGHI, J.M.:
Estimation of direct genetic and maternal effects for production traits of Iranian Holstein cows using different animal models. *Pak. J. Biol. Sci.* **9**(4) (2006), 636-640
- GRASER, H.U.; SMITH, S.P.; TIER, B.:
A derivative-free approach for estimating variance components in animal models by restricted maximum likelihood. *J. Anim. Sci.* **64** (1987), 1362-1370
- HARDER, B.; BENNEWITZ, J.; REINSCH, N.; MAYER, M.; KALM, E.:
Effect of missing sire information on genetic evaluation. *Arch. Tierz., Dummerstorf* **48** (2005), 219-232
- HENDERSON, C.R.:
Use of all relatives in intraherd prediction of breeding values and producing abilities. *J. Dairy Sci.* **58** (1975), 1910-1916
- HERRENDÖRFER, G.; TUCHSCHERER, A.; DIETL, G.; TUCHSCHERER, M.:
The robustness of BLUP and EBLUP (German language). *Arch. Tierz., Dummerstorf* **42** (1999), 303-318
- IM, S.; FERNANDO, R.L.; GIANOLA, D.:
Likelihood inferences in animal breeding under selection: A missing-data theory view point. *Genet. Sel. Evol.* **21** (1989), 399-414
- KHATTAB, A.S.; ATIL, H.; BADAWY, L.:
Variances of direct and maternal genetic effects for milk yield and age at first calving in a herd of Friesian cattle in Egypt. *Arch. Tierz., Dummerstorf* **48** (2005), 24-31
- KREJČOVÁ, H.; MIELENZ, N.; PŘIBYL, J.; SCHÜLER, L.:
Comparison of breeding values for daily gains of bulls estimated with multi-trait and random regression models. *Arch. Tierz., Dummerstorf* **50** (2007b), 147-154
- KREJČOVÁ, H.; MIELENZ, N.; PŘIBYL, J.; SCHÜLER, L.:
Estimation of genetic parameters for daily gain of bulls with multi-trait and random regression models. *Arch. Tierz., Dummerstorf* **50** (2007a), 37-46
- LUTAYA, E.; MISZTAL, I.; BERTRAND, J.K.; MABRY, J.W.:
Inbreeding in populations with incomplete pedigrees. *J. Anim. Breed. Genet.* **116** (1999), 475-480
- MEYER, K.:
DFREML: Programs to estimate variance components by restricted maximum likelihood using a derivative-free algorithm. User notes, Version 3.0β. University of New England, Armidale, Australia (1998)
- NDLOVU, P.:
Classical and bootstrap estimates of heritability of milk yield in Zimbabwean Holstein cows. *J. Dairy Sci.* **76** (1993), 2013-2024
- NILFOROOSHAN, M.A.; EDRISS, M.A.:
Comparison of Holstein bull semen sources on milk traits in Isfahan province in Iran. *Arch. Tierz., Dummerstorf* **50** (2007), 71-83
- SANDERS, K.; BENNEWITZ, J.; KALM, E.:
Wrong and missing sire information affects genetic gain in the Angeln dairy cattle population. *J. Dairy Sci.* **89** (2006), 315-321
- SAS, 1997:
SAS/STAT User's guide, Version 6.12, SAS Institute, Cary, NC, USA.
- WILCOX, C.J.; WEBB, D.W.; DELORENZA, M.A.:
Genetic improvement of dairy cattle. Institute of Food and Agricultural Sciences, University of Florida, Document DS094. <http://edis.ifas.ufl.edu/DS094> (1992)

Received: 2007-03-12

Accepted: 2008-01-28

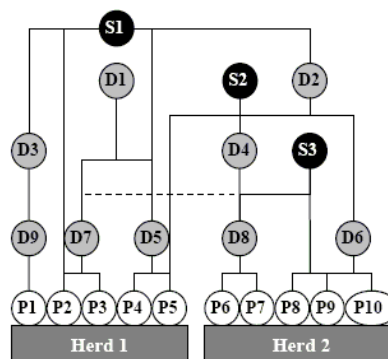
Corresponding Author
 Prof. MOHAMMAD ALI EDRISS
 Department of Animal Science
 Isfahan University of Technology
 ISFAHAN, IRAN. 8415683111

email: edriss@cc.iut.ac.ir

On sabbatical leave:
 Div. Genetics and Genomics
 Roslin Institute (Edinburgh)
 Midlothian EH25 9PS
 U.K.

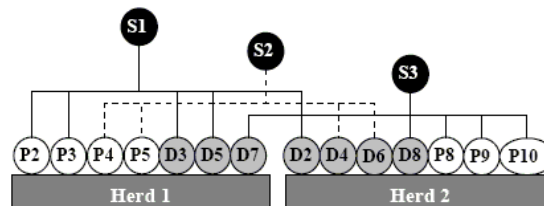
Appendix (EDRISS et al.)

Diagram 1: A sample of a pedigree structure (containing 4 generations, 3 sires/grandsires, 9 dams/grand dams and 10 progenies in 2 herds)[§] in an animal model for a sex-limited trait (no parent for sires) (Eine Probe einer Zucht-Struktur (4 Erzeugungen, 3 sires/grandsires, 9 dams/grand Verdammungen und 10 Nachkommen in herds)§ 2 in einem Tiermodell für ein Geschlecht-begrenztes Merkmal (kein Elternteil enthalten für Herren))



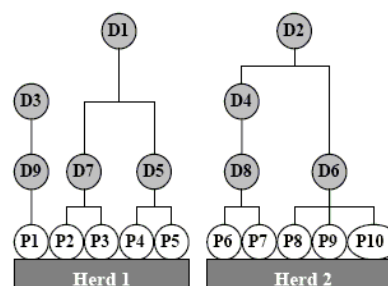
§ S= Sire, D= Dam and P= Progeny. Dotted line connects S3 to D7.

Diagram 2: Considering Diagram 1, the following relationships can be remained by a sire model (dam missed)[§] (Diagramm 1 betrachtend, können die folgenden Verhältnisse durch ein Herrmodell (die Verdammung vermißt worden) geblieben werden)



§ S= Sire, D= Dam (here, D2-D8 act only as progeny) and P= Progeny.

Diagram 3: Considering Diagram 1, the following relationships can be remained by a dam model (sire missed)[§] (Diagramm 1 betrachtend, können die folgenden Verhältnisse durch ein Verdammung Modell (der Herr vermißt worden) geblieben werden)



§ D= Dam and P= Progeny.