

Estimates of Genetic and Environmental (Co)Variances for First Lactation Milk Yield, Survival, and Calving Interval

M. C. DONG and L. D. VAN VLECK
Department of Animal Science
Cornell University
Ithaca, NY 14853

ABSTRACT

Variance and covariance components for milk yield, survival to second freshening, and calving interval in first lactation were estimated by REML with the expectation and maximization algorithm for an animal model which included herd-year-season effects. Cows without calving interval but with milk yield were included. Each of the four data sets of 15 herds included about 3000 Holstein cows. Relationships across herds were ignored to enable inversion of the coefficient matrix of mixed model equations. Quadratics and their expectations were accumulated herd by herd. Heritability of milk yield (.32) agrees with reports by same methods. Heritabilities of survival (.11) and calving interval (.15) are slightly larger and genetic correlations smaller than results from different methods of estimation. Genetic correlation between milk yield and calving interval (.09) indicates genetic ability to produce more milk is slightly associated with decreased fertility.

INTRODUCTION

Fertility is the second most important cause of culling. Of all disposals, cows culled due to fertility problems accounted for about 28% in western Europe (2, 3, 23), 16% in the US, and 27% in Israel (6). An important concern is the genetic relationship between production and reproduction. Because management practices and breeding policies have been changing, further study on (co)variance structure between production and reproduction seems desirable

now that new statistical methodology and necessary computing power are available.

Results from previous studies suggest the relationship between milk production and fertility is complex. Although some studies from field data show an antagonistic relationship between milk yield and fertility in dairy cows (4, 5, 9, 21, 25, 28) and a complementary relationship among measures of fertility from breeding heifers and measures of their first lactation production (8, 17, 18), "more knowledge is needed before the genetic relationships are known well enough to determine what, if any, changes are needed in selection practice" (7). Calving interval (CI), influenced by all fertility traits, can be used to estimate the genetic covariance between production and reproduction. To eliminate bias caused by selection, REML for an animal model should be applied to data on which selection decisions were made (19, 20, 26).

In most cases, data provided by dairy herds have been subjected to varying intensities of sequential selection. Methods to estimate (co)variance components between traits (10, 24) have required animals to have both traits recorded. If selection on one or more traits has occurred, estimates are biased. To eliminate bias caused by selection, data on which selection decisions are based should be included in analysis and methods such as maximum likelihood and REML (22) should be used (13). The objective of this study was to estimate (co)variances among milk yield, survival, and calving interval using REML.

MATERIALS AND METHODS

In this study, REML by the expectation and maximization algorithm (1) was used to estimate (co)variances among milk yield, calving interval, and survival to second freshening. Calving interval, the sum of days open and gestation length, is an overall measure of fer-

Received April 22, 1988.
Accepted August 31, 1988.

tility for cows that conceive and calve again (16). Survival to second freshening is an indicator of selection in first lactation for various reasons including poor fertility. The survival trait was included in an attempt to obtain estimates of covariances between first lactation milk yield and calving interval that are not conditional on the cow freshening a second time.

The analyses were:

1) Estimation of genetic (co)variances among milk yield, CI, and survival in first lactation using records of cows that had milk yield and survival records but did not have CI. Survival was included in the model to obtain unconditional estimates, because a cow has a CI conditional on whether she has a second freshening date.

2) Estimation of genetic (co)variances among milk, CI, and survival in first lactation using only selected records, i.e., after eliminating cows without CI. Survival was not included in the model because only cows that survived have records. The reason for this analysis was to compare results from the conditional and unconditional models. The same data sets were used as by Dijkhuizen (1).

Milk yield is 305-d, mature equivalent (ME) milk in the first lactation. Calving interval is the number of days between first and second freshening dates. Survival is defined as one for cows with a second freshening date, zero otherwise.

Model

An animal model accounting for relationships was used.

$$y = X\beta + Z\alpha + \epsilon \quad [1]$$

where:

- y is the vector of observations,
- β is a vector of fixed herd-year-season effects,
- α is a vector of additive genetic values of individual animals for the traits,
- ϵ is a vector of residual effects,
- X and Z are known incidence matrices that assign fixed and random effects to records in y.

Expectations and variances are defined as:

$$E \begin{bmatrix} \alpha \\ \epsilon \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \text{Var} \begin{bmatrix} \alpha \\ \epsilon \end{bmatrix} = \begin{bmatrix} G & 0 \\ 0 & R \end{bmatrix}$$

The mixed model equations are:

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{bmatrix} \begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix} \quad [2]$$

where b and a are estimates of β and α . Let n denote the number of animals and t the number of traits. Data are ordered traits within animals and missing observations on CI are accounted for by zero columns in X and Z. R is block diagonal with n blocks (R_i) of order t. Each R_i is derived from R_0 , the $t \times t$ residual (co)variance matrix. For an animal with no missing records, $R_i = R_0$. For an animal with no calving interval observed, elements of the last row and column in R_0 are set to zero.

Instead of R^{-1} , R^- is used in the mixed model equations because for an animal with missing values the diagonal blocks, R_i^- , are derived from R_i with zeroed rows and columns.

In this study, the only missing records are for CI. Thus, there are two types of R_i . For an animal with complete records:

$$R_i^- = R_i^{-1} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{12} & r_{22} & r_{23} \\ r_{13} & r_{23} & r_{33} \end{bmatrix}^{-1}$$

where 1, 2, and 3 identify milk, survival, and calving interval, respectively. For an animal with missing CI:

$$R_i^- = \begin{bmatrix} r_{11} & r_{12} & 0 \\ r_{12} & r_{22} & 0 \\ 0 & 0 & 0 \end{bmatrix}^{-1}$$

$G^{-1} = A^{-1} * G_0^{-1}$, where * denotes the direct product operation, A is the numerator relationship matrix, and G_0 is the additive genetic (co)variance matrix of order t.

The method applied to estimate \mathbf{R}_0 and \mathbf{G}_0 was REML as described by Henderson (13, 14). The method is the same as iterated MIVQUE (15) except at each round of iteration, expectations are taken under pretense $\tilde{\mathbf{G}} = \mathbf{G}$ and $\tilde{\mathbf{R}} = \mathbf{R}$, where $\tilde{\mathbf{G}}$ and $\tilde{\mathbf{R}}$ are solutions from the prior round for the additive genetic and residual (co)variance matrices \mathbf{G}_0 and \mathbf{R}_0 , respectively [see (27) for a description of the computing procedure].

The assumption to ignore relationships across herds made by Swalve and Van Vleck (27) was made in this study so that solving [2] was possible by accumulating the quadratics for REML herd by herd. A further reduction in the number of equations was obtained by not setting up equations for base animals, although including them to compute the relationship matrix. Base animals were those that did not have records but created relationships among animals with records [see, e.g., (27)].

Data

Data chosen for this study were from cows freshening in 1970 through 1984. Thus, there could have been 14 (years) \times 2 (seasons) \times 3 (traits) = 84 fixed effects for each herd. For an average herd size of 200 cows (range from 180 to 220), the order of a herd block in coefficient matrix of [2] is 684, which took 800 s to invert and 100 s to set up equations, write elements of inverse to tape, and compute quadratic forms with an IBM-4381. Thus, for a data set consisting of 15 herds, 3 h and 45 min were required for one round of iteration. Evidence by Swalve and Van Vleck (27) using REML suggested that at least 18 to 20 rounds of iteration were necessary to obtain reasonably converged estimates of (co)variances. Thus, in this study about 70 h would have been needed to obtain estimates for one data set. Therefore, another simplification was made: daughter-dam relationships were ignored. Then when sires of cows with same base sires are grouped together, \mathbf{A}^{-1} for each herd is block diagonal corresponding to a base sire which results in block diagonal form of $\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1}$ in [2]. Then a generalized inverse of [2] for a herd was carried out by applying rules for inversion of a partitioned matrix [e.g., (12)] taking advantage of the block diagonality of $\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1}$ within herd. By this method, each round needed 2 h

and 10 min. The disadvantage is that the estimates may be biased somewhat downward dependent on the number of dam relationships in a herd. Henderson (11) showed numerically that if sire relationships were ignored, the residual variance increased. Sorensen and Kennedy (26) studied estimation of genetic variances from unselected and selected populations using MIVQUE by simulation. Their results showed that for an animal model with complete relationship matrix among breeding values, estimates of additive genetic variances were unbiased in both unselected and selected populations, but when a sire model where progeny were assumed to be related only through their sires was used, estimates of additive genetic variances were biased downward a little in an unselected population and considerably in a selected population. Sampling variances of estimated variances from the animal model were also smaller than those from the sire model.

The original data set consisted of records on 864,181 artificially sired New York Holstein cows, of which 663,643 (76.8%) had second lactations that first freshened between January 1, 1970 and December 31, 1984. The reason for not using data after 1984 was to give all cows opportunity to have a complete CI in first lactation.

The first criterion for editing the data was that a cow had a milk yield >1818 kg in first lactation no matter whether she had a valid CI. A total of 752,867 cows met the criteria.

After eliminating cows with condition codes affecting records (sample >75 d, sick or injured, aborted, nurse cows), there were 590,552 cows left. To avoid skipped lactations and recording errors, CI was restricted to between 300 and 600 d. Only 1.5% of cows had CI less than 300 d (.33%) or greater than 600 d (1.17%). Finally, 581,347 cows remained for study. A computationally manageable data set had to be selected, as will be described later.

Likelihood functions can be improved by increasing the data per herd. However, herds with many cows also increase computing time because of the need to invert the coefficient matrix. The need for large herds must be balanced against computing time. Further, 30 herds with a small number of sires and a large number of daughters per sire were selected.

TABLE 1. Structure of the two data sets for study 1.

Traits	Data set	
	1	2
Milk yield		
Number of records	3063	3077
Mean, kg	7886	7233
Survival		
Number of records	3063	3077
Mean (probability)	.80	.85
Calving interval		
Number of records	2456	2606
Mean, d	382.4	387.2
Average cows per herd	204	205
Average equations per herd	697	699

Two data sets were chosen with each set containing 15 herds. Herds were selected from the range of herd sizes of 190 to 220 for cows with milk yield in first lactation. A summary of the data structure for the two data sets is given in Table 1.

Total rounds of iteration were 23 for data set 1 and 20 rounds for data set 2. Initial estimates for iteration for data set 2 were those from round 21 for data set 1. Starting variances for data set 1 were derived from an among and within sires analysis. Covariances in G_0 and R_0 were set to zero as starting values.

RESULTS AND DISCUSSION

Estimation of Relationships Among Milk Yield, Survival, and Calving Interval in First Lactation

Table 2 shows estimates of parameters obtained from round 23 in data set 1 and round 20 from data set 2. Estimates in both data sets are similar. Heritability of milk yield is in general agreement with results by other authors for different Holstein populations. Swalve and Van Vleck (27) reported heritabilities in first lactation from two data sets to be .33 and .34, which are slightly larger than the results from this study. Their study included herds with higher production than herds in this study. Heritabilities of survival and CI are slightly larger than results in literature using different methods of estimation. This may be due to

TABLE 2. Estimated heritabilities and genetic and phenotypic correlations for milk yield, survival, and calving interval in first lactation from two data sets.

Estimates	Data set	
	1	2
Heritabilities		
Milk	.33	.28
Survival	.11	.12
Calving interval	.17	.14
Genetic correlations		
Milk, survival	.09	.11
Milk, calving interval	.10	.08
Survival, calving interval	-.05	-.06
Phenotypic correlations		
Milk, survival	.28	.28
Milk, calving interval	.28	.23
Survival, calving interval	.04	.07
Residual variances		
Milk, (kg/1364) ²	.56	.75
Survival, 0 or 2.5	.82	.70
Calving interval, (d/6) ²	.54	.62

using REML with an animal model, which may be less affected by selection in comparison with other methods and other models.

All phenotypic correlations are greater than genetic correlations. Phenotypic correlations are in good agreement with results by Miller et al. (21) using paternal half-sib analysis. Phenotypic parameters are similar for the two studies, although the data sets were from populations in two different time periods. However, genetic correlations by Miller et al. (21) were large, ranging from .43 to .65 for milk production and CI, from .53 to .77 for milk production and survival, and from .24 to 1.02 for survival and CI. Large genetic correlations between milk production and CI and milk production and survival were explained as due to confounding between pleiotropy and selection for milk production. The range of genetic correlations between survival and CI was great. The largest estimate was out of the parameter space, suggesting that confounding with some fixed effects was not eliminated in their method. REML may have reduced the effects of selection.

From the current study, the small genetic correlation between milk yield and CI indicates

TABLE 3. Structure of the two data sets when cows without calving interval were excluded.

Traits	Data set	
	1	2
Milk yield		
Number of records	2456	2606
Mean, kg	8032	7912
Calving interval		
Number of records	2456	2606
Mean, d	382.4	387.2
Average number of cows/herd	164	173
Average number of equations/herd	382	400

selection on milk production would result in only a slight increase in CI. In other words, increasing genetically the ability to produce more milk would decrease fertility only slightly. The small heritability of CI suggests that fertility problems should be approached by improvement in management.

The result that the phenotypic correlation is larger than the genetic correlation between milk production and survival suggests that high producing cows survive longer largely due to non-genetic factors, one of which is the fact that high producers may be given more chance for survival. Genetic and phenotypic correlations

between first CI and survival are not significantly different from zero in agreement with Wilcox et al. (29).

Estimation of Relationships Between Milk Yield and Calving Interval in First Lactation, When Cows Without Calving Interval Were Excluded

Two data sets used in this analysis were the same as in the previous section, except that cows without CI were excluded; 2456 cows (80%) were left in data set 1 and 2606 cows (85%) were left in data set 2. Data structures for the two data sets are in Table 3. In comparison with the means of milk yield for the two data sets using all data shown in Table 1, means of milk yield for the same two data sets using selected data were larger (Table 3), indicating cows discarded in first lactation tended to be low producers. Variances shown in Tables 2 and 4 also suggest that selection of cows to have a second lactation had occurred.

Table 4 shows estimated parameters obtained at round 18 for both data sets. In comparison with the results shown in Table 2, heritability of milk yield was slightly larger. Thus, when selected data were used, estimated heritability may have been biased upward although the sample sizes are too small to provide a definite conclusion. However, heritability of CI remained almost the same, indicating little difference between conditional and unconditional heritability for CI. There also was only a

TABLE 4. Estimated heritabilities and genetic and phenotypic correlations for milk yield and calving interval in first lactation from two data sets using only records of cows with calving interval.

Estimates	Data set 1	% of All data ¹	Data set 2	% of All data ²
Heritabilities				
Milk	.35	106	.31	111
Calving interval	.17	102	.14	98
Correlations				
Genetic	.16	151	.07	96
Phenotypic	.27	97	.24	101
Residual variance				
Milk (kg/1364) ²	.46	82	.58	78
Calving interval (day/60) ²	.53	99	.62	99

¹ Estimates for data set 1 in Table 2.

² Estimates for data set 2 in Table 2.

small difference in phenotypic correlations from selected data and unselected data. The genetic correlation increased in data set 1 using selected data and decreased a little in data set 2 using selected data.

Estimates of residual variance of milk yield decreased after loss of 20% of records in data set 1 and 15% in data set 2. Estimates for genetic variance decreased about 10% and residual variance decreased about 20% in both data sets. However, estimates for genetic covariance increased 40% in data set 1 and decreased 10% in data set 2, whereas estimates for residual covariance decreased about 20% for both data sets. Estimates for genetic and residual variances of CI, however, remained almost the same. Because survival was dropped out in these two data sets, estimates for and with CI are conditional results. Comparison of the two studies showed little difference between estimates of the conditional variance and the unconditional variance for calving interval such that from data set 1; standardized estimates of the unconditional genetic and residual variances were .1084 and .5392 and estimates of the conditional genetic and residual variances were .1094 and .5329.

Results from the two selected data sets suggest when selected data sets are used, that, as is well known, estimates for both genetic and residual variances can be biased downward. Heritability, however, from selected data may be biased upward if the estimated residual variance decreases proportionally more than estimated genetic variance. Including survival in the model does not seem important for estimating phenotypic correlation between first lactation production and CI.

CONCLUSIONS

Heritability of milk yield in first lactation is in agreement with reports by the same methods. Heritabilities of survival and CI in first lactation are slightly larger and genetic correlations are smaller than the results in the literature that did not use REML with an animal model. The small genetic correlation between milk yield and CI indicates the genetic ability to produce more milk is only slightly associated with decreased fertility. Therefore, based on these limited data genetic selection for milk would not seem to be a major cause of

fertility problems. The small heritability of CI suggests fertility problems can best be approached by improvement in management.

When selected data sets (made up only of cows with both milk yield and calving interval) were used to estimate genetic (co)variances for milk yield and calving interval, estimates of both genetic and residual variances decreased, but estimates of residual variances decreased relatively more than genetic variances. Thus, heritability seemed to be biased upward from analysis of the selected data. Genetic correlations from selected data were also larger than from unselected data, although there was little difference in phenotypic correlations between selected and unselected data. The small differences between conditional and unconditional results indicate inclusion of survival in the model to obtain unconditional results for CI was not important.

ACKNOWLEDGMENTS

This research was supported in part by a grant from Eastern AI Cooperative, Inc., and from the Institute of Animal Science, Chinese Academy of Agricultural Science and the Chinese Ministry of Agriculture, Animal Husbandry and Fishery.

REFERENCES

- 1 Dempster, A. P., N. M. Laird, and D. B. Rubin. 1977. Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. B.* 39:1.
- 2 Dijkhuizen, A. A. 1980. De economische betekenis van gezondheidsstoornissen bij melkvee. I. Voortijdige afvoer. Publ. No. 4., Afd. Agric. Econ. Vakgroep Zootech., Fac. Diergeneeskunde, Utrecht.
- 3 Drees, B. 1982. Genetisch-statistische Auswertungen zur Fruchtbarkeit von Milchrindern in Schleswig-Holstein. Thesis, Christian Albrechts Univ., Kiel.
- 4 Eckles, C. H. 1929. A study of breeding records of dairy herds. *Minnesota Agric. Exp. Stn. Bull.* 258, St. Paul.
- 5 Everett, R. W., D. V. Armstrong, and L. J. Boyd. 1966. Genetic relationships between production and breeding efficiency. *J. Dairy Sci.* 49:879.
- 6 Freeman, A. E. 1984. Secondary traits: sire evaluation and reproductive complex. *J. Dairy Sci.* 67: 449.
- 7 Freeman, A. E., and P. M. VanRaden. 1984. Optimum systems and information needed for genetic gain through artificial breeding. Proc. Natl. Invitational Workshop Genet. Improvement of Dairy Cattle, Milwaukee, WI.
- 8 Hansen, L. B., A. E. Freeman, and P. J. Berger. 1983. Variances, repeatability, and age adjustments

- of yield and fertility in dairy cattle. *J. Dairy Sci.* 66:281.
- 9 Hansen, L. B., A. E. Freeman, and P. J. Berger. 1983. Yield and fertility relationships in dairy cattle. *J. Dairy Sci.* 66:293.
 - 10 Henderson, C. R. 1953. Estimation of variance and covariance components. *Biometrics* 9:226.
 - 11 Henderson, C. R. 1975. Best linear unbiased estimation and prediction under a selection model. *Biometrics* 31:423.
 - 12 Henderson, C. R. 1976. A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. *Biometrics* 32:69.
 - 13 Henderson, C. R. 1984. Application of linear models in animal breeding. Univ. Guelph, Ont., Can.
 - 14 Henderson, C. R. 1984. Estimation of variances and covariances under multiple trait models. *J. Dairy Sci.* 67:1581.
 - 15 Hocking, R. R., and M. H. Kutner. 1975. Some analytical and numerical comparisons of estimators for the mixed AOV model. *Biometrics* 31:19.
 - 16 Majjala, K. 1964. Fertility as a breeding problem in artificially bred populations of dairy cattle. I. Registration and heritability of female fertility. *Ann. Agric. Fenn.* 3(Suppl. 1):1.
 - 17 Majjala, K. 1976. Possibilities of improving fertility in cattle by selection. *World Rev. Anim. Prod.* 12:69.
 - 18 Metz, J.H.M., and R. D. Politiek. 1970. Fertility and milk production in Dutch Friesian cattle. *Neth. J. Agric. Sci.* 18:72.
 - 19 Meyer, K. 1985. Genetic parameters for dairy production of Australian black and white cows. *Livest. Prod. Sci.* 12:205.
 - 20 Meyer, K., and R. Thompson. 1984. Bias in variance and covariance component estimators due to selection on a correlated trait. *Z. Tierz. Zuchtungsbiol.* 101:33.
 - 21 Miller, P. D., L. D. Van Vleck, and C. R. Henderson. 1967. Relationships among herd life, milk production, and calving interval. *J. Dairy Sci.* 50:1283.
 - 22 Patterson, H. D., and R. Thompson. 1971. Recovery of interblock information when block sizes are unequal. *Biometrika* 54:545.
 - 23 Philipsson, J. 1981. Genetic aspects of female fertility in dairy cattle. *Livest. Prod. Sci.* 8:307.
 - 24 Searle, S. R., and T. R. Rounsaville. 1974. A note on estimating covariance components. *Am. Stat.* 28:67.
 - 25 Seykora, A. J., and B. T. McDaniel. 1983. Heritabilities and correlations of lactation yields and fertility for Holsteins. *J. Dairy Sci.* 66:1486.
 - 26 Sorensen, D. A., and B. W. Kennedy. 1984. Estimation of genetic variances from unselected and selected populations. *J. Anim. Sci.* 59:1213.
 - 27 Swalve, H. H., and L. D. Van Vleck. 1987. Estimation of genetic (co)variances for milk yield in first three lactations using an animal model and restricted maximum likelihood. *J. Dairy Sci.* 70:842.
 - 28 Whitmore, H. J., W. J. Tyler, and L. E. Casida. 1974. Effects of early postpartum breeding in dairy cattle. *J. Anim. Sci.* 38:339.
 - 29 Wilcox, C. J., K. O. Pfau, and J. W. Bartlett. 1957. An investigation of inheritance of female reproductive performance and longevity, and their interrelationships within a Holstein-Friesian herd. *J. Dairy Sci.* 40:942.