Improvement in Accuracy Using Records Lacking Sire Information in the Animal Model.

by

Changhee Do

Dissertation submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Animal Science(Dairy)

APPROVED:

_____
William E. Vinson, Chairman and
Department Head

_____
Ronald E. Pearson

_____
David R. Notter

_____
Bennet G. Cassell

_____
Klaus Hinkelmann

Blacksburg, Virginia

Improvement in Accuracy Using Records Lacking Sire Information in the Animal Model.

by

Changhee Do

William E. Vinson, Chairman and Department Head

Animal Science(Dairy)

(ABSTRACT)

Four alternative methods were examined with computer simulated data to improve accuracy of animal model genetic evaluations by including records lacking sire identification. Methods 1 and 2 assumed genetic values of cows missing sire identity were population and management group average, respectively. Methods 3 and 4 accounted for genetic values through producing abilities estimated as random and fixed effects, respectively. Correlations between true and estimated management group effects and breeding values of cows and sires were used as measures of estimation accuracy. Alternative methods were examined to determine 1) optimum, minimum management group size, 2) increases in estimation accuracy of alternative methods relative to the conventional method of discarding records lacking sire identity, 3) the effects on accuracy of missing sire identity for lower true breeding value sires, and 4) the potential to use different alternative methods in herds of varying size, proportion of cows sire identified, and level of variation.

Management group effects were estimated more accurately as minimum management group size increased (3 to 6 to 9), but breeding values were less accurate. Accuracies of alternative methods slightly exceeded those of the conventional method for all estimated effects and all minimum group sizes.

Accuracies of alternative and conventional methods were compared in 60 population with 250 sires and averages of 11,139 cows with 23,849 records. Alternative

methods were always more accurate than the conventional method for estimating group effects. Methods 1 and 3 were uniformly more accurate in estimating breeding values of cows, and estimated breeding values of sires more accurately in 55 and 54, respectively, of 60 populations. Increases in accuracy were largest for method 3, but small for all methods.

Intentionally omitting identity for daughters of sires with low breeding value reduced accuracy of estimation for breeding values but not for group effects. However, alternative methods were more accurate than the conventional method. Alternative methods were relatively most accurate for estimating breeding values in small herds having high variance and low proportions of sire identified cows. Method 3 had uniformly highest accuracy but method 1 often was similar with less computing cost.

Acknowledgements

The author wishes to express sincere appreciation to Dr. William E. Vinson for his guidance, knowledge, support, and encouragement during his work toward Doctorate degree. Special thanks are due to Dr. Ron E. Pearson, Dr. Bennet G. Cassell, Dr. David R. Notter, and Dr. Klaus Hinkelmann for serving on his graduate committee and for their help in the preparation of this dissertation, and also, his appreciation goes to Dr. Ina Hoeschele for intensive training during one and one-half years of his work with her.

The author wishes to acknowledge the contribution of his wife, Jeeyoung, whose patience and understanding made completion of this document possible. Encouragement and support of his family in Korea also made it possible to accomplish difficult tasks in the United States.

Appreciation is also due to the departmental computer programmers and fellow genetic graduate students for their programming tips and for helping him to understand the American dairy industry.

Table of Contents

List of Illustrations

List of Tables

INTRODUCTION

Genetic improvement in dairy cattle populations requires appropriate genetic analysis of performance records. Genetic evaluation methods have been developed and refined through several decades to estimate true genetic merit for production and non-yield traits, independent of environmental influences. Due to the magnitude of the dairy cattle genetic evaluation problem, full models of inheritance historically have not been employed in genetic evaluation. Recently, the animal model (97) for additive genetic effects and the dominance model (35) for dominance and epistatic effects have become computationally feasible due to the development of more efficient computing algorithms and the extension of computing capacity. The benefit of genetic evaluation using more complete models is increased accuracy of selection and an acceleration of genetic improvement.

Genetic improvement in dairy cattle is determined mostly by bull selection, even though accurate genetic evaluation facilitates genetic improvement through bull dam selection. Successful production of sires for future AI use is based on mating genetically elite cows and bulls to obtain young bulls for progeny testing. Such pedigree selected young bulls are then progeny tested by obtaining production records of their daughters. Progeny tests of young sires are required to provide estimated breeding values of sufficient accuracy for choosing future AI bulls. However, the number of usable daughters per young sire is limited to about 25 on average at the first proof. Improved accuracy of estimating breeding values of young sires at an early stage of progeny testing would allow more young sires to be culled earlier, thus reducing the cost of progeny testing with the same selection intensity, or increasing selection intensity through testing more young

bulls at the same cost. Also, cows currently are selected as dams of future young sires based on estimated breeding values of relatively low accuracy. Improvement in accuracy of estimating breeding values of dams of future young sires would allow sampling of young sires of higher genetic merit and, thus, increase the rate of genetic progress.

Many daughter records are eliminated in current dairy cattle genetic evaluation. The major cause of the elimination is missing sire identification on records, which causes rejection of about 2 of 5 otherwise acceptable records (95). Reasons for missing sire identification are not clearly identifiable, but probably include both intentional and unintentional (i.e., poor record keeping) omissions. The latter, likely representing a large majority of instances, should result in a normal distribution of breeding values of cows lacking sire information. The former, however, would likely result in a distribution skewed toward lower breeding values. The proportions of intentional and unintentional omission are unknown, but instances of the former have not been reported in the literature.

Genetic evaluation in dairy cattle has emphasized accurate evaluation of sires due to the potential wide use of sires through AI and consequent impact on genetic improvement. Since daughter records lacking sire identification contribute nothing to the evaluation of her sire, such records have been discarded from genetic evaluation procedures. Also, estimated genetic values of cows with unknown sires are very unreliable, since genetic merits of cows are transmitted from their sires and dams, and accuracy of estimated genetic values of dams is usually low. When use of the records lacking parent's identity is considered, the genetic values of those records must be accounted for in a way to avoid biases. In the USDA animal model, genetic values of unknown parents are assumed equal to the average genetic value of individuals in the "genetic group" to which they belong. Criteria for assigning animals to genetic groups are somewhat arbitrary (33, 74, 76, 92). Most commonly used criteria are birth date and geographic region,

which account for differences in average genetic value due to genetic trend and genetic difference by region, perhaps due to regional differences in breeding goals or practices. Accuracy of the estimated genetic group mean is directly related to the size of the genetic group which is dependent upon flexible criteria. In contrast, the accuracy with which the estimated group mean reflects the genetic value of an individual in the group varies inversely to within-group variation and hence, likely also to group size. The genetic values of records lacking sire or parent identity can be considered analogous, since genetic groups could be as small as management groups, or as large as all animals born in the past several years in a geographical region. Another way to account for genetic values of cows lacking parent identity is to estimate producing ability from cow's records only, without considering sire and dam genetic merit.

Production records are used in genetic evaluation to estimate average environmental as well as genetic effects since separation of genetic and environmental contributions to performance is a principle requirement in genetic evaluation. The current U.S. Department of Agriculture (USDA) implementation of the animal model defines for each production record, a management group effect which is used to adjust records for differences in environmental influences. Five production records with sire identity are considered as a minimum for accurate estimation of the average group effect. In the event five records are not available, group dimensions are expanded to an arbitrarily set maximum group size in an attempt meet this minimum. Initially, management groups are defined by herd, year, 2-month season, registry status (registered or grade) and parity (first or later). For groups lacking five records, season is initially expanded to four months. Registry distinction is dropped next, followed by an expansion of seasons to six months. For groups still failing to contain five records, the parity distinction is dropped and the number of required records reduced to three. Lastly, seasons are expanded by two month intervals to 12 months and two records are required.

Discarding records lacking sire information aggravates the shortage of group mates and requires the combining of adjacent groups, thereby reducing the homogeneity of the group effect for records in the same group. Loss of records may have particularly undesirable consequences on accuracy of genetic evaluation in breeds other than Holstein. Records lacking sire identification could be potentially used to construct environmental groups which include larger numbers of group mates and/or are more homogeneous with respect to environmental factors affecting milk yield. Consequently, improved estimation of breeding value may be expected. The goal of this study was to find alternative methods to improve accuracy of genetic evaluation utilizing records which lack sire identity.

REVIEW OF LITERATURE

Since no work has been done directly related to this study, this section will describe contemporary groups and genetic evaluation procedures in general in order to orient the reader to the basis of this study. The approach of this section is limited to support only this study, and thus, some sources of variation in milk yield and characteristics of genetic evaluation procedures used in dairy cattle are not reviewed.

Environmental Effects on Milk Yield

To obtain accurate and unbiased estimates of breeding values of animals, the extraneous variance that contributes to biases must be taken into account. Sources of environmental variance in milk yield of dairy cattle include herd, year-season of freshening, age, parity, registry status, special treatment of certain cows, lactation length, days-open, as well as interactions among those effects, and so on. Some of these factors are not easily measured or may be confounded with genetic effects such as preferential treatment of certain cows and registry status.

Considerable work (3, 8, 27, 77) has shown that the single largest source of variation in milk yield is due to herd-year-season of freshening. Variance due to herd differences in climatic conditions and levels of feeding and management has been reported to range from 25 to 50% of total phenotypic variation in milk yield. Variance due to the interaction of herd with month and or year-season has been reported to explain from 3 to 5% of total variation and variance due to seasonal difference of calving month alone was 1.4% (3). The impracticality of developing correction factors for individual herd-

year-seasons has led to the use of genetic evaluation methods which express performance relative to that of other individuals in the same herd-year-season (2, 6, 15, 63, 87) or which adjust for herd-year-season differences simultaneously with estimation of genetic merit (20, 73).

Adjustment for age effects by multiplicative correction factors generally is used in genetic evaluations of dairy cattle under the assumption that effects of age at calving on milk yield are environmental in nature. Factors for age correction ideally are estimated so as to minimize confounding with level of management, genetic trend, region and selection (11, 28, 41, 43, 52). Variances due to age, and interaction of age and herds have accounted for 17 and 3% of total variance for milk yield, but after correction by multiplicative age factors, these values are .7 and 3.4% (83). After correction for age effects, the effect of parity and the interaction of age and parity have been found to be small and non-significant (42, 53). Interaction of age with herd and month of freshening have been estimated to account for three to four percent of total variation in milk yield. Effects of age on milk yield have been found to differ across months or season of calving (53) but not across herds (82) and thus, age correction factors within season have been suggested. Herd-sire interaction has accounted for 2% of the variance in milk yield, but 16% of the variance in age at calving (37), suggesting that adjustment for age within sire-herd may desirable, even though computationally prohibitive.

Parity is an environmental effect to the degree that yields in different lactations are genetically the same trait. However, parity (after age) has been considered in genetic evaluation procedures to account for the genetic effects of female selection (6, 97). Average bias in evaluations of sires from female selection has been found to be small (2, 13, 94) but effects on evaluations of young, relative to older, sires may be especially undesirable. In practice, parity has been distinguished as "first" and "later" (6, 97), which

assumes most selection occurs on first lactation performance, and the two parities placed in separate herdmate (6) or management (97) groups.

Powell and Norman (70) reported environmental and slight genetic differences between registered and grade Holstein cows. Registered cows were in herds with environment that supported high milk yield more often than were grades, and tended to be managed preferentially to grades in the same herd. Sire-identified Holstein grade cows had higher genetic merit for milk yield than did registered cows (68). Registered cows outperformed grades as 2-year old cows for milk yield in the same herd year, but registered superiority lessened or disappeared in later lactations because grades were culled more intensely (59). Registered and grade cows have been placed in separate management groups in genetic evaluation to account for this difference (97).

Evaluation Methods With Sire Models

The wide spread use of artificial insemination (AI) in dairy cattle has resulted in heavy research emphasis on developing accurate procedures for genetic evaluation of sires. Since the early 1950's, several methods have been used, either explicitly or implicitly, that could be called a "sire model" (or a variant, the "sire-maternal grand sire model") in which the only specified genetic effects were associated with sires (or sires and maternal grand sires) of animals supplying records (17, 72).

A principal requirement of any sire evaluation method is that it remove at least a major proportion of the differences among sires in environmental factors which affect daughter milk yield (20). Correction of records for differences in age and length of lactation leaves differences due to variation in climate, feeding , management, etc., which are many times larger than the differences in sire genetic merit to be estimated (3). Lacking the computing ability to estimate simultaneously environmental effects of indi-

vidual herds, years, and seasons, sire evaluation procedures have adopted the procedure of expressing daughter production of a sire relative to (as a deviation from) the average performance of other cows subjected to similar environmental conditions.

The daughter-dam (DD) comparison, one of the earliest methods of sire evaluation, compared the production of daughters of a sire with that of his mates. In the absence of cow selection, DD provides correction for phenotypic merit of mates as well as herd, but not year or season, environmental effects. The large effects on milk yield of differences among sires in year and season environmental effects was a major reason for the abandonment of the DD comparison in favor of comparisons to individuals producing more nearly contemporaneously in time.

National sire evaluations by herdmate comparison (HC) procedures were implemented in the U.S. in 1962. Adjustments for differences among sires in management of daughters was accomplished by expressing yield for daughters of a given sire relative to yield for daughters of other sires (herdmates) calving in the same herd, year, and season (2, 15, 87). Contemporary comparison (CC) procedures were similar except that herdmates were required to be of the same parity as daughters. The CC procedure was used less widely due to the low frequency of contemporary as compared to herdmate records (44). The HC procedure did not attempt any adjustment for genetic merit of mates of sires but was modified to include a correction for differences in average genetic merit of herdmates (7) as

$$PD = R[D - HM + 0.1(HM - BA)] \qquad [1.1]$$

where PD is the estimated sire genetic merit (transmitting ability) or Predicted Difference, R is theoretical accuracy of the estimate or Repeatability, D-HM is the average difference between daughter and herdmate production, and BA is breed average production. The term 0.1(HM-BA) adjusted the evaluation for differences in genetic merit

among herdmates of a sire under assumption that, on average, 10% of phenotypic differences among herdmate averages were genetic in origin.

Comparison among alternative sire evaluation procedures (54, 65) and estimates of potential bias in HC methods (45, 46, 58) resulted in their replacement by the Modified Contemporary Comparison (MCC) in 1974. The major changes in methodology which characterized the MCC (6, 7, 63) were 1) a yield comparison involving "mostly" cows of the same (first or later) parity, 2) an adjustment for differences among sires in genetic merit of "mostly" contemporary herdmates of their daughters (now called modified contemporaries), 3) an adjustment for differing impact of genetic trend on young and older sires by regressing evaluations to a genetic group average defined by pedigree genetic merit, rather than to breed average, and 4) establishment of a fixed, known genetic base. Algebraically, the MCC can be represented (7) as

$$PD = R(D - MCA + SMC) + (1 - R)GA = R(MCD) + (1 - R)GA \qquad [1.2]$$

where D-MCA is the average difference between yield of daughters and modified contemporaries, SMC is average genetic merit of sires of modified contemporaries (computed iteratively), MCD is daughter yield relative to modified contemporaries and adjusted for genetic merit of contemporary sires, GA is a group average MCD determined from the sire's pedigree index estimate of genetic merit (66), and other terms are as indicated previously. Prior to 1983, GA was the average MCD of sires with similar pedigree index (PI), computed as

$$PI = .5(\text{sire PD}) + .25(\text{MGS PD}) \qquad [1.3]$$

Subsequently, GA was replaced by an estimate of ancestor merit (AM) such that the MCC estimate of sire genetic merit (96) was

$$PD = R(MCD) + (1 - R)AM \qquad [1.4]$$

with AM (96) computed from either PI (equation [1.3]) or, where an estimate of the dam's genetic merit (Cow Index or CI) was available, from

$$GI = .5(sire \ PD) + .5(dam \ CI) \qquad [1.5]$$

as

$$AM = \alpha + d + GI + b \ B \qquad [1.6]$$

or

$$AM = \alpha + d + PI + b \ B \qquad [1.7]$$

where d is an average difference associated with AI vs. non-AI sampling, B is relative birth date of the sire in days, and b is the regression of MCD for the pedigree group on relative birth date. Bulls were assigned to groups based on breed, birth year and pedigree information (GI or PI) in computing AM.

Best Linear Unbiased Prediction (BLUP) procedures developed by Henderson (20, 21) also used sire and sire-maternal grand sire (72) models in the Northeast AI sire comparison. BLUP sire model procedures incorporated features of selection index, such as maximization of the expected mean of animals with equal information chosen by truncation selection and the correlation between predictor and predictand, with a feature of least squares procedures, namely, minimization of mean square errors of estimated fixed effects. Both estimation of random genetic merit and adjustments for environmental effects were carried out simultaneously. BLUP sire model procedures incorporating additive genetic relationships (22, 23) were introduced by Henderson (25) to partially account for genetic trend and genetic differences among sub-populations, and

thereby reduce the necessity for genetic groups and increase accuracy of breeding value estimation.

Best Linear Unbiased Prediction with an Animal Model

Unlike BLUP sire models, animal models include genetic relationships for males as well as females. Increased genetic information from relatives greatly reduces the necessity for genetic groups and increases accuracy of genetic evaluation in comparison to the sire model. However, numbers of equations to be solved are remarkably increased in the animal model. Suppose milk yield is described as

$$y = X\beta + Za + Zp + e \qquad [1.8]$$

where β, a, p and e are contemporary group, additive genetic, permanent environmental and unexplained effects on milk yield, respectively, and X and Z are the incidence matrices for corresponding effects. Approximate comparisons between BLUP animal model evaluations and those by HC or MCC can be made by examining estimates of genetic merit. With the BLUP animal model, after adjustment for contemporary group effects on milk yield the effect of the ith cow on milk yield is

$$\pi_i = a_i + p_i \qquad [1.9]$$

with the expected value of the cow effect partitioned as

$$\pi_i = .5a_s + .5a_d + p_i \qquad [1.10]$$

where a_s and a_d are additive genetic effects of sire and dam, respectively, and p_i is a permanent environmental effect. Since an animal model contains the genetic value of the

animal, neither p_i nor all or parts of a_d are in the error term of equation [1.8]. After absorbing contemporary group effects from equation [1.8] and ignoring permanent environmental effects for illustration, the mixed model equations are

$$\begin{bmatrix} Z'MZ + A^c\alpha & A^{cs}\alpha \\ A^{sc}\alpha & A^s\alpha \end{bmatrix} \begin{bmatrix} a_c \\ a_s \end{bmatrix} = \begin{bmatrix} Z'My \\ 0 \end{bmatrix} \qquad [1.11]$$

where α is the ratio of error variance to additive genetic variance, $M = X(X'X)^- X'$ and the inverse of the relationship matrix for cows ($[A_c]_{k \times k}$), sires ($[A_s]_{(n-k) \times (n-k)}$) and cow-sires ($[A_{cs}]_{k \times (n-k)}$) is

$$A^{-1} = \begin{bmatrix} A_c & A_{cs} \\ A_{sc} & A_s \end{bmatrix}^{-1} = \begin{bmatrix} A^c & A^{cs} \\ A^{sc} & A^s \end{bmatrix}_{(n \times n)} \qquad [1.12]$$

where n, k and n-k are the numbers of animals, cows and sires to be evaluated, respectively. An estimated breeding value of the ith cow is expressed from equation [1.11] as

$$\hat{a}_i = \frac{1}{(Z'MZ + A^c\alpha)_{ii}} \left[Z'My_i - \sum_{\substack{j=1 \\ j \neq i}}^{k} (Z'MZ + A^c\alpha)_{ij}\hat{a}_j - \sum_{\substack{j=k+1 \\ j \neq i}}^{n} A^{cs}_{ij}\alpha\hat{a}_j \right] \qquad [1.13]$$

Let

$$w_2 YD = \left[Z'My_i - \sum_{\substack{j=1 \\ j \neq i}}^{k} Z'MZ_{ij}\hat{a}_j \right] / (Z'MZ + A^c\alpha)_{ii} \qquad [1.14]$$

where YD is similar to an average yield deviation of the cow's records (86) from average yield of management group mates. Substituting elements of the relationship matrix inverse gives as the estimated breeding value of the cow,

$$\hat{a}_i = w_2 YD + \frac{-\alpha}{(Z'MZ + A^c\alpha)_{ii}} \left[\hat{a}_d A^c_{i,d} + \hat{a}_s A^{cs}_{i,s} + \sum_{mates} \hat{a}_m A^{cs}_{i,m} + \sum_{progeny} \hat{a}_p A^{-1}_{i,p} \right]$$

where d, s, m and p are column numbers for dam, sire, mates and progeny, respectively, $A^c_{i,d}$ and $A^\alpha_{i,s}$ are -1, $A^\alpha_{i,m}$ is (number of progeny$=np$)/2, $A^{-1}_{i,p}$ is -1, and a_p and a_m are additive genetic effects of the cow's progeny and mates, respectively. Let $w_1 = 2\alpha/(Z'MZ + A^c\alpha)_{ii}$ and $w_3 = \alpha/\{2np(Z'MZ + A^c\alpha)_{ii}\}$. Then the estimated breeding value of the cow (86) is

$$\hat{a}_i = w_1(.5\hat{a}_s + .5\hat{a}_d) + w_2 YD + w_3\overline{(2\hat{a}_p - \hat{a}_m)} \tag{1.15}$$

The w's are weights for each term. Genetic information from all relatives and mates as well as the cow's own records are used in computing estimated genetic merit. For an estimated breeding value of a bull with both parents known, equation [1.11] gives

$$\left[\hat{a}_d A^{sc}_{i,d} + \sum_{mates} \hat{a}_m A^{sc}_{i,m} + \sum_{progeny} \hat{a}_p A^{-1}_{i,p} + \hat{a}_s A^s_{i,s} + \hat{a}_i A^s_{i,i} \right] \alpha = 0$$

where $A^s_{i,i}$ is 2 since both parents of the bull are known. Then, the estimated breeding value of the bull can be similarly written as

$$\hat{a}_i = (.5\hat{a}_s + .5\hat{a}_d) + .25\overline{(2\hat{a}_p - \hat{a}_m)} \tag{1.16}$$

Since bulls do not have measurements on milk yield, the breeding values are estimated through parents, progeny, and mates.

The animal model evaluation for bulls can be contrasted with an analogous evaluation using a sire model. Suppose the model is

$$y = X\beta + Zs + e^* \qquad [1.17]$$

where s is a vector of sire effects. For a sire model, $.5a_d + p_i$ in equation [1.10] is included in e^* in equation [1.17], and thus, the error term becomes larger. An estimated breeding value of a bull in the sire model (84) can be written analogous to [1.16] and [1.4] as

$$\hat{a}_i = w_4(.5\hat{a}_s) + w_5(MCD) \qquad [1.18]$$

where w_4 and w_5 are functions of "effective number of daughters" (89, 100). Comparison of [1.16] and [1.18] demonstrates the more complete specification of genetic effects in the former. The term $w_5(MCD)$ is analogous to $.25\overline{(2\hat{a}_p - \hat{a}_m)}$, but lacks adjustment for genetic merit of mates. Also, MCD as in equation [1.4] adjusts only for genetic merit of sires of contemporaries while $\hat{a}_p$ and $\hat{a}_m$ are adjusted for genetic merit of contemporaries themselves (86). Likewise, genetic merit of both sire and dam are considered in [1.16] rather than $\hat{a}_s$ only. Lastly, animal model evaluations are simultaneous for cows and bulls. With the sire model, evaluations of cows are computed following sire evaluation and are approximate.

Animal model evaluations can be compared to evaluations from a sire-maternal grand sire model by further partitioning equation [1.10].

$$\pi_i = .5a_s + .25a_{mgs} + .25a_{mgd} + p_i \qquad [1.19]$$

where a_{mgs} and a_{mgd} are additive genetic effects of the maternal grand sire and grand dam, respectively. Estimated genetic merit of sires and maternal grand sires were used in the MCC procedure, to calculate ancestor merit, and in BLUP sire-maternal grand sire models to adjust directly evaluations of bulls (72). Suppose equation [1.17] represents a sire-maternal grand sire model with the vector s including breeding values of maternal grand sires in addition to sires. Then e^{*} in [1.17] contains $.25a_{mgd} + p_i$ in equation [1.19], which is smaller fraction of genetic effects than in the sire model. In the sire-maternal grand sire model estimated breeding value of a bull can be written analogous to equation [1.18] as

$$\hat{a}_i = w_6(.5\hat{a}_s + .25\hat{a}_{mgs}) + w_7(\text{MCD}) \hspace{3cm} [1.20]$$

The sire-maternal grand sire model has one additional source of genetic variation ($.25a_{mgs}$) removed from error relative to the sire model. Therefore the sire-maternal grand sire model is theoretically more accurate than the sire model.

Accurate genetic evaluation of cows is especially important for selecting cows to serve as bull dams. Cow evaluations from a sire model typically have been of the form (69)

$$\hat{a}_i = w_8(.5\hat{a}_s + .5\hat{a}_d) + w_9(\text{MCD}) \hspace{3cm} [1.21]$$

Both parents provide genetic information to the cow evaluation, but progeny information in equation [1.21] is completely ignored. In addition, evaluations of cows under a sire model are not simultaneous with sire evaluation, nor have they typically been iterated solutions for $\hat{a}_i$ and $\hat{a}_d$.

The USDA Animal Model

The theory underlying genetic evaluation using an animal model was developed decades ago (19) and has been used for beef cattle genetic evaluation (73). Recently (July, 1989) the animal model was implemented by USDA for genetic evaluation in large dairy cattle populations. This was made possible by the development of improved computing facilities and algorithms over the past several years. The USDA animal model can be described in matrix notation (97) as

$$y = X\beta + Za + ZQg + Pp + Cc + e \qquad\qquad [1.22]$$

where y represents a vector of records; β, a, g, p, and c are vectors of effects for management groups, the random portion of additive genetic merit, the unknown-parent group portion of genetic merit, permanent environment, and herd-sire interaction, respectively; X, Z, ZQ, P, and C are incidence matrices for these effects; and e is error. Vectors a, p, c, and e are treated as random effects which are mutually uncorrelated, and β and g are treated as fixed effects. The management and genetic groups defined in the model will be explained in detail subsequently. The relationship of animals to unknown ancestor groups (76, 92) is described by the matrices Q and A, where A is the additive genetic relationship matrix among all animals in a. The herd-sire effect in the animal model is included to account for environmental correlations among daughters of a sire in a common herd (97, 98).

The relationship matrix A has a major role in mixed model genetic evaluation using the animal model. The relationship matrix specifies all known genetic relationships among animals and takes into account relationships in the mixed model equations. Suppose a relationship matrix for five animals is

$$A = \begin{bmatrix} 1.00 & 0 & .50 & 0 & .25 \\ 0 & 1.00 & .50 & 0 & .25 \\ .50 & .50 & 1.00 & 0 & .50 \\ 0 & 0 & 0 & 1.00 & .50 \\ .25 & .25 & .50 & .50 & 1.00 \end{bmatrix}$$

Animals 1 and 2, corresponding to rows 1 and 2, are sire and dam of cow 3, and animals 4 and 3 are sire and dam of cow 5. Breeding values can be expressed using relationships specified in A, as

$$a = \begin{bmatrix} a_1 \\ a_2 \\ a_3 = \underline{.5a_1 + .5a_2} + m_3 \\ a_4 \\ a_5 = \underline{.5a_4 + .25a_1 + .25a_2} + m_5 \end{bmatrix}$$

where, for cow 3, a_1 and a_2 are breeding values of sire and dam, respectively, and m_3 represents the random Mendelian sampling contribution to the breeding value of cow 3 due to segregation of parental genotypes during gametogenesis. In a sire model, $.5a_2$ and $.25a_2$ in the underlined information from parents are not utilized, and genetic evaluations for cows are not computed directly. The following equation for cow 3 from [1.15] clearly shows the flow of information

$$\hat{a}_3 = w_1(.5\hat{a}_1 + .5\hat{a}_2) + w_2 YD + w_3(2\hat{a}_5 - \hat{a}_4)$$

where YD is yield deviation of cow 3, which is defined as the cow's yield average adjusted for all effects other than genetic merit and error, and w's are weights for each

source of information. Information from sire, dam and progeny is used for estimating breeding value of cow 3 through the relationship matrix and cows and bulls are evaluated simultaneously (34). Information from siblings flows through parents, and the animal is credited with the effects of genes which are transmitted to, and measured in the performance of progeny (86). In effect, w_2YD and $w_3(2\hat{a}_5 - \hat{a}_4)$ are used to estimate m_3 and potentially to correct for any differences between true and estimated values for a_1 and a_2. Proper weighting according to genetic relationship assures that an individual's superiority due to inheritance from one parent is not improperly credited to the other parent. For example the superiority of cow 5 due to bull 4 is subtracted from the evaluation of cow 3 as $w_3(2\hat{a}_5 - \hat{a}_4)$.

Genetic groups have been considered in dairy cattle genetic evaluation to account for genetic trends and differences of sub-populations (7, 97). Arbitrary definition of genetic groups may include sex, breed, birth date, birth place, AI stud and so on. The USDA animal model includes genetic groups which are defined for unknown parents (97). For a sire or dam to be assigned to an unknown parent group, three criteria must be met: 1) both parents are assigned to unknown parent groups, 2) only one progeny, and 3) no yield record. Unknown parent groups are assigned separately for each selection path, sire of sire, dam of sire, sire of cow, and dam of cow to allow for differing average genetic values for different paths. Within each path, animals are assigned by birth year with the number of years in a group determined to equalize group size and avoid groups too small for accurate estimation (97).

Procedures for estimating genetic group effects are complicated when relationships among animals are used. From the example given previously, if the six parents of animals 1, 2 and 4 represent unknown parents, the breeding values of unknown parents appears in the breeding values of progeny with s_1, s_2, s_4, d_1, d_2 and d_4 as

$$
a = \begin{bmatrix} a_1 = .5s_1 + .5d_1 + m_1 \\ a_2 = .5s_2 + .5d_2 + m_2 \\ a_3 = .5a_1 + .5a_2 + m_3 \\ a_4 = .5s_4 + .5d_4 + m_4 \\ a_5 = .5a_4 + .5a_3 + m_5 \end{bmatrix}
$$

The vector of breeding values can be rewritten as

$$
a = \begin{bmatrix} .5 & 0 & 0 & .5 & 0 & 0 & : & 0 & 0 & 0 & 0 & 0 \\ 0 & .5 & 0 & 0 & .5 & 0 & : & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & : & .5 & .5 & 0 & 0 & 0 \\ 0 & 0 & .5 & 0 & 0 & .5 & : & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & : & 0 & 0 & .5 & .5 & 0 \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \\ s_4 \\ d_1 \\ d_2 \\ d_4 \\ a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \end{bmatrix} + \begin{bmatrix} m_1 \\ m_2 \\ m_3 \\ m_4 \\ m_5 \end{bmatrix}
$$

Then, with matrix notation (76)

$$
\begin{aligned}
a &= [P_b : P]\begin{bmatrix} a_b \\ a \end{bmatrix} + m \\
&= P_b a_b + Pa + m \\
&= (I - P)^{-1}(P_b a_b + m)
\end{aligned}
\tag{[1.23]}
$$

where a_b and a represent vectors of breeding values for unknown parents and animals, respectively, and the matrix $[P_b : P]$ relates progeny to parents. Each row of the matrix contains two .5 elements. Since the expectation of Mendelian sampling is a vector of zeros, the expectation of equation [1.23] (76) is

$$
E(a) = (I - P)^{-1}P_b E(a_b)
\tag{[1.24]}
$$

The expectation of genetic values of unknown parents is the genetic group effect of unknown parents such that $E(a_b) = Q_b g$ where Q_b is a matrix relating unknown parents to their unknown parent genetic groups. In the example given, if genetic groups are defined by generation, the following would represent the expectation of genetic effects of unknown parents.

$$E(a_b) = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} g_1 \\ g_2 \end{bmatrix}$$

where each row corresponds to s_1, s_2, s_4, d_1, d_2 and d_4. Let $Q = (I - P)^{-1} P_b Q_b$, and then $E(a) = Qg$. The term $(I - P)^{-1}$ can be written in a power series as

$$(I - P)^{-1} = I + P^1 + P^2 + P^3 + \dots$$

where the P matrices represent an animal's relationship to ancestral generation according to the power of the matrix; e.g., one for parents, two for grand parents, three for great grand parents etc.. In the example given,

$$P^1 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ .5 & .5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & .5 & .5 & 0 \end{bmatrix} \quad \text{and} \quad P^2 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ .25 & .25 & 0 & 0 & 0 \end{bmatrix}$$

Because in this example there is no ancestor beyond grand parents among identified animals, all matrices except for P^1 and P^2 are null. Adding matrices I, P^1, and P^2 gives

$$(I - P)^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ .5 & .5 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ .25 & .25 & .5 & .5 & 1 \end{bmatrix}$$

After multiplications of $(I - P)^{-1}$, P_b and Q_b for $Q = (I - P)^{-1} P_b Q_b$,

$$E(a) = Qg$$

$$= \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ .5 & .5 \end{bmatrix} \begin{bmatrix} g_1 \\ g_2 \end{bmatrix}$$

Therefore

$$u = \begin{bmatrix} g_1 + a_1 \\ g_1 + a_2 \\ g_1 + a_3 \\ g_2 + a_4 \\ .5g_1 + .5g_2 + a_5 \end{bmatrix}$$

where u is the vector of total (random a_i + fixed g_i) genetic merits of animals (86). The genetic effect of each animal therefore includes random as well as fixed genetic effects, the latter coming from unknown parent groups whose effects are related to animals through the relationship matrix. Since parents of cow 5 are animals 3 and 4 in the example, the genetic group effect of cow 5 is the average effect of groups 1 and 2. As a more general illustration suppose permanent environment and herd-sire effects in equation [1.22] are ignored. Then the mixed model equations including genetic group effects (76) are

$$\begin{bmatrix} Q'Z'MZQ & Q'Z'MZ \\ Z'MZQ & Z'MZ + A^{-1}\alpha \end{bmatrix} \begin{bmatrix} \hat{g} \\ \hat{a} \end{bmatrix} = \begin{bmatrix} Q'Z'My \\ Z'My \end{bmatrix} \qquad [1.26]$$

After solving the equation [1.26], total genetic merits are estimated as

$$\hat{u} = Q\hat{g} + \hat{a} \qquad [1.26]$$

The genetic group effects of unknown ancestors ($\hat{g}$) are treated as fixed effects and ($\hat{a}$) are treated as random effects in the mixed animal model.

Estimated group effects account for genetic trends over time and regions, if the definition for grouping is appropriate (21, 22). Further, the animal model accounts for selection, if all data on which selections are based are used, the relationship matrix is complete, and the model used is correct (12, 23). Since the relationship matrix reflects all pathways of selection or non-random mating, potential biases from selection and non-random mating are avoided (23).

Some potential problems, however, remain in the animal model. If selection is applied based on breeding values estimated with a relationship matrix, a combination of individual and family selection is practiced. This is a good reason to use the relationship matrix, since combined selection is always superior to individual or family selection (9). The effectiveness of alternative selection schemes is determined by the intraclass correlations among breeding values and phenotypic values of family members (40). If "environmental" (actually, environmental plus non-additive genetic) correlations exist between relatives, the distributional assumptions, $N(0, I\sigma_p^2)$ and $N(0, I\sigma_e^2)$ for environment and error effects in the additive animal model are not valid, which might imply existence of dominance or epistatic effects. In such a situation the weight given to various sources of information by the relationship matrix may be incorrect since phenotypes are assumed to be related by additive genetic relationships only. Thus, common environmental effects in a family may reduce the accuracy of breeding values estimated using the relationship matrix, and a reduction of selection efficiency may result (9).

Preferential treatment of selected cows and heterogeneous variances within herds can markedly affect genetic evaluations of cows and young sires with few progeny, and these problems may be accentuated by the use of relationship matrix. Biases due to preferential treatment might easily accumulate through relationships of maternal pedigree (grand-dam $\rightarrow$ dam $\rightarrow$ daughter), and performance records can be exaggerated

(positive or negative relative herd mean) in high variance herds (20, 90). These problems can be reduced in a sire model using the maternal grand sire instead of dam.

Management Groups in the Animal Model

Management groups are defined in genetic evaluation procedures to account for and reduce differential effects on performance records due to variation in management and environment associated especially with herd and time differences. The definition of contemporary groups for dairy cattle might include herd-year-season of freshening, lactation number, registered or non-registered status, and special treatment (such as a hormone treatment) among others. Management groups defined (97) in the USDA animal model require a minimum 5 lactation records and are classified by year-month of calving (2-month periods), registry status (registered or grade), and parity (first or later) in a herd and year. The maximum possible number of management groups in a year for a herd would be 24, necessitating a minimum of 120 lactations. Many herds are well short of this number and require the combining of groups by relaxing requirements of grouping. Requirements for grouping are relaxed in order as follows to the point required to obtain a required number of contemporary group records (97).

- season is increased to 4 months.
- registered and grade distinction removed.
- season is increased to 6 months.
- parity distinction removed and minimum size is reduced to 3.
- season is increased to 12 months in increments of 2 months, minimum size reduced to 2.

The Effect of Combining Groups

The grouping strategy for management and environment can affect both accuracy and bias of genetic evaluation (89). Numerous groups may provide a model close to the true model, but can result in small numbers of records per group, a danger of ill-behaved estimation, and increased prediction error variance. However, the likelihood of bias from failure to adjust properly for group conditions is increased by combining true groups, since it reduces homogeneity among group mates. Thus, combining groups may reduce prediction error variance and increase bias, while less combining may reduce bias and increase prediction error variance. In practice, mean square error, which includes bias squared plus prediction error variance, may be a more meaningful measure of optimum group structure than bias or prediction error variance alone. Thus, a balance between bias and prediction error variance should be carefully considered to minimize mean square error (89).

Bias

In dairy cattle genetic evaluation, records can be classified for multi-way analysis of fixed effects such as registry status, herd, year, season and parity. It is conjectured that differences in milk production records can be well explained by an analysis of multi-way classification when numbers of observations in subclasses are large enough to estimate fixed effects accurately. This has not been, however, applied in practice. Rather the concept of management groups (one way classification for the fixed effects) has been employed in genetic evaluation with most models including the current USDA animal model. For accurate estimation of group effects, the USDA animal model has established the somewhat arbitrary minimum of five records per group (97). As an illustration

of bias from combining groups, suppose a vector of 5 records from 5 animals in 4 groups, which can be described by either a two-way or a one-way model.

$$y = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} p_3 \\ p_4 \end{bmatrix} + e$$

$$y = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} + e$$

For example, p_1 and p_2 might represent different herds while p_3 and p_4 correspond different years. Then, $b_1 = p_1 + p_3$, $b_2 = p_1 + p_4$, $b_3 = p_2 + p_3$, and $b_4 = p_2 + p_4$. If one were forced to combine management groups b_1 with b_2, and b_3 with b_4, the resulting model would ignore the effects of p_3 and p_4; the model formed as a consequence of combining groups is

$$y = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} + e$$

The second incident matrix times $[p_3 \ p_4]'$ of the first model is ignored by combining groups. It was shown by Henderson (24) that ignoring certain fixed effects actually present in the model leads to biased estimators. As an example, suppose the expectation of a vector of random genetic effects is a vector of zeros such that $E(\hat{u}) = 0$ and the true model is

$$y = X_1\beta_1 + X_2\beta_2 + Zu + e$$

$X_2\beta_2$ could represent the vector ignored in the first model. When we ignore $X_2\beta_2$, the estimates of random genetic effects are biased as

$$E(\hat{u}) = (C_{12}{}'X_1{}'X_2 + C_{22}Z'X_2)\beta_2$$

where C_{12} and C_{22} are the 1,2[th] and 2,2[th] block matrices of the inverse of the coefficient matrix, respectively, from the mixed model equations for β_1 and u. Since the expectation of the estimated genetic effects is a function of $X_2\beta_2$, combining of groups potentially increases the bias in estimations of genetic effects.

Prediction Error Variance

The mixed model equations written with simple notation after absorbing fixed effects into additive genetic effects, is a useful tool to illustrate the effect of additional group mates. Let n_{ij} denote the number of records of the ith animal in the jth management group and y_{ij} denote the sum of records of those animals. The mixed model equations with q levels of a random effect (20) are

$$
\begin{bmatrix}
n_{1\bullet} - \sum\limits_j \dfrac{(n_{1j})^2}{n_{\bullet j}} + \alpha & -\sum\limits_j \dfrac{(n_{1j})(n_{2j})}{n_{\bullet j}} & \cdots\cdots & -\sum\limits_j \dfrac{(n_{1j})(n_{qj})}{n_{\bullet j}} \\
 & n_{2\bullet} - \sum\limits_j \dfrac{(n_{2j})^2}{n_{\bullet j}} + \alpha & \cdots\cdots & -\sum\limits_j \dfrac{(n_{2j})(n_{qj})}{n_{\bullet j}} \\
 & & \cdots\cdots & \\
 & & \cdots\cdots & \\
\text{symmetric} & & & n_{q\bullet} - \sum\limits_j \dfrac{(n_{qj})^2}{n_{\bullet j}} + \alpha
\end{bmatrix}
\begin{bmatrix}
\hat{a}_1 \\
\hat{a}_2 \\
\cdot \\
\cdot \\
\cdot \\
\hat{a}_q
\end{bmatrix}
=
\begin{bmatrix}
y_{1\bullet} - \sum\limits_j n_{1j}\overline{y}_{\bullet j} \\
y_{2\bullet} - \sum\limits_j n_{2j}\overline{y}_{\bullet j} \\
\cdot \\
\cdot \\
\cdot \\
y_{q\bullet} - \sum\limits_j n_{qj}\overline{y}_{\bullet j}
\end{bmatrix}
$$

where α is σ^2_e/σ^2_a. The inverse of the coefficient matrix times σ^2_e is the prediction error variance-covariance matrix for animals 1 through q. The element $[n_{i\bullet} - \sum\limits_j (n_{ij})^2/n_{\bullet j} + \alpha]$ in the coefficient matrix is called the "effective number of daughters" (89, 100) in the sire

model and is the figure most related to accuracy of genetic evaluation of sires. A corresponding value in the animal model is referred as "daughter equivalents" (86). Adding more management group mates in the jth group provides larger $n_{\bullet j}$. Consequently the daughter equivalents will be larger due to the subtraction of a smaller number from $n_{j\bullet}$. Large values of daughter equivalents give small prediction error variance, since the diagonal elements of the inverse matrix become smaller for larger diagonal elements of the coefficient matrix.

MATERIALS AND METHODS

The distribution of herd size in populations of dairy cattle is not normal, but rather is skewed toward small herd size relative to a normal distribution. Consequently it often is difficult to obtain sufficient numbers of records in subclasses of non-genetic factors which affect milk yield. Accuracy in estimation of breeding values is closely related to accuracy of estimating environmental group effects (89). Records lacking sire identification have limited value in estimating breeding values but may be useful for improving estimation of group effects. In this context, objectives of this study were to determine the validity of using data abandoned due to missing sire identification and to develop alternative methods for using such records in animal model genetic evaluation.

Simulation

Simulation of Milk Yield

Simulations were done for milk yield of dairy cows for 11 years. The first six years of simulation was used as a foundation to prime the system. Breeding values of cows and bulls were simulated and random mating was practiced under the assumption of no inbreeding. An average of 25 daughters per sire was intended. Records of a cow were not simulated after the fifth lactation even if she had offspring afterward. For the purpose of simulation, herd-sire effects were excluded from the model. The model for simulating a mature equivalent milk record is as follows

$$y_{ijkl} = \mu + g_{ij} + a_k + p_k + e_{ijkl} \qquad [2.1]$$

where μ is the population mean, g_{ij} is the effect of the jth management group in the ith herd, a_k is the breeding value of the kth cow, p_k is the permanent environmental effect of the kth cow, and e_{ijkl} is the temporary environmental effect on the lth record of the kth cow.

The jth management group effect in the ith herd contains the temporary herd effect (effect of a 2-month season in a specific year) and the permanent effect of the ith herd. The equation for simulating g_{ij} was

$$g_{ij} = z_i\sqrt{\sigma^2_{ph}} + z_{ij}\sqrt{\sigma^2_{th}} \qquad [2.2]$$

where z_i and z_{ij} are uncorrelated standard normal deviates, and σ^2_{ph} and σ^2_{th} are variances of permanent and temporary herd effects, respectively. Each record of cows in a herd was randomly assigned to a 2-month season in a given year. A randomly chosen z_i was common to all records in a given herd and a randomly chosen z_{ij} was common to all records in a given herd-year-season. To simulate the breeding value of the kth animal with pedigree information from both parents, the following equation was used

$$a_k = .5a_s + .5a_d + z_k\sqrt{.5\sigma^2_a} \qquad [2.3]$$

where a_s and a_d are the breeding values for sire and dam, respectively of animal k, z_k is a standard normal deviate, and σ^2_a is variance of breeding values. Mendelian sampling of animal k was simulated by $z_k\sqrt{.5\sigma^2_a}$. The equation for the permanent environmental effect on all records of animal k was defined as follows

$$p_k = z_k'\sqrt{\sigma^2_p}$$

where σ^2_p is variance of permanent environmental effects of cows, and z_k' is a different standard normal deviate from z_k for animal k. Permanent environmental effects may include dominance and epistatic effects in addition to true environmental effects. Since herd-sire interaction was excluded in the simulation, the sum of breeding value and permanent environmental effect of a cow is her producing ability. Finally the equation for the temporary environmental effect on the lth record of animal k is

$$e_{ijkl} = z_{ijkl}\sqrt{\sigma^2_e}$$

where σ^2_e represents variance of temporary environmental effects of records. This effect contains variation from any source not included elsewhere in the model. Parameters in the simulation were chosen to create variability similar to that reported for milk yield. Effects in the model (excepting μ) were assumed to be normally distributed with mean 0 and dispersion parameters given below in kg (47).

Parameters	μ	$\sqrt{\sigma^2_a}$	$\sqrt{\sigma^2_{ph}}$	$\sqrt{\sigma^2_{th}}$	$\sqrt{\sigma^2_p}$	$\sqrt{\sigma^2_e}$
Given values	6,804	528.4	780.6	552.0	528.4	747.5
Estimated values		531.4	767.6	559.4	530.1	748.0

In general, the estimated values from 27,536 simulated records were approximately equal to the given parameter values.

Simulation of Populations

Each cow was randomly assigned to a bull for mating with the probability of use being equal for all bulls. The probability that each mating produced a female offspring which completed a first lactation was .343 (62). Culling (including death) of animals was simulated following the completion of each lactation. As a proportion of animals

completing first lactation, culling rates (62) were for each lactation: 1st, .217; 2nd, .208; 3rd, .187; 4th, .138; 5th, .086; 6th, .071; 7th, .040; 8th, .031; 9th, .029. Probabilistic decisions to have given matings produce offspring completing first lactation and to cull after each lactation were made by uniform random numbers.

Initial herd sizes and proportions of sire-identified cows were either assigned or randomly generated and used for different purposes. Assigned initial herd sizes of 40, 70 and 100 were used for investigating the influence of herd size on accuracy of genetic evaluation, as shown in d) of Table 1. Initial herd sizes also were generated using random numbers from a generalized distribution developed from the actual distribution of herd sizes in Virginia DHI herds in 1990 (Figure 1) excluding herd larger than 300 (2.2%), and used to provide variation in herd size for investigating impacts of minimum management group sizes on accuracy of genetic evaluation (a in Table 1), probability of increasing accuracy by alternative methods (b in Table 1) and influence of intentionally omitted identification of lower breeding value sires (c in Table 1). Likewise, proportions of sire-identified cows for each herd were either assigned (20, 50 and 80% in d of Table 1) or randomly generated and used for different purposes. Randomly generated proportions of sire-identified cows used random numbers from a generalized distribution developed from the actual distribution of sire identified percentage in Virginia DHI herds excluding extremely low and high proportions ($< .1$ and $> .9$) shown in Figure 2. The population in Figure 2 was designated the "medium" level of sire identification. "Low" and "high" levels of sire identification had the same general distributions but with mean proportions shifted $\pm .1$ from the medium level population. The correlation between herd size and proportion of sire-identified cows estimated from Virginia DHI herds was low (-.025) and ignored in simulation. After a sire-identified proportion for each herd was generated by selecting a random number from the distribution, sire identity was deleted from the required proportion of cows by selecting a

Table 1. Basic statistics in simulations.

Simulations	Herds	Sires	Herd size	Sire ID level	Replications	Records	Cows	Cows with sire ID		Daughters per sire	Minimum group size
								Number	Percent		
a)	60	250	general	low	10	24,111	11,263	5,327	47.3%	21.31	3, 6, 9
	60	250	general	medium	10	24,604	11,480	6,408	55.8%	25.63	3, 6, 9
	60	250	general	high	10	24,370	11,355	7,735	68.1%	30.94	3, 6, 9
				average		24,362	11,366	6,490	57.1%	25.96	
b)	60	250	general	low	20	23,219	10,857	5,141	47.4%	21.06	5
	60	250	general	medium	20	24,254	11,318	6,535	57.7%	26.22	5
	60	250	general	high	20	24,053	11,233	7,629	67.9%	30.27	5
				average		23,849	11,139	6,399	57.4%	25.63	
c)	60	250	general	high	5	24,917	11,644	7,566	65.0%	30.93	5
	Second Phase of Deletion						0%	6,186	53.1%	30.43	5
							10%	6,085	52.3%	30.23	5
							20%	6,046	51.9%		
d)	80	230	40	80%	3, 3*	15,561	7,265	5,779	79.8%	25.21	5
	80	140	40	50%	3, 3	15,857	7,395	3,652	49.4%	26.09	5
	80	50	40	20%	3, 3	15,389	7,177	1,406	19.6%	28.12	5
	60	300	70	80%	3, 3	20,535	9,567	7,650	80.0%	25.50	5
	60	180	70	50%	3, 3	20,269	9,643	4,783	49.6%	26.57	5
	60	70	70	20%	3, 3	20,691	9,673	1,917	19.8%	27.39	5
	40	300	100	80%	3, 3	19,661	9,152	7,306	79.8%	24.35	5
	40	180	100	50%	3, 3	19,608	9,158	4,571	49.9%	25.39	5
	40	70	100	20%	3, 3	19,743	9,214	1,856	20.1%	26.51	5

a), b), c), and d) describe simulations in the determination of minimum group sizes, the probability of increasing accuracy, the influence of omitted lower breeding value sires, and the performance under assigned conditions on herds, respectively.
* Low (552kg) and high (733kg) levels for standard deviation of temporary herd effects were simulated for three replications each.
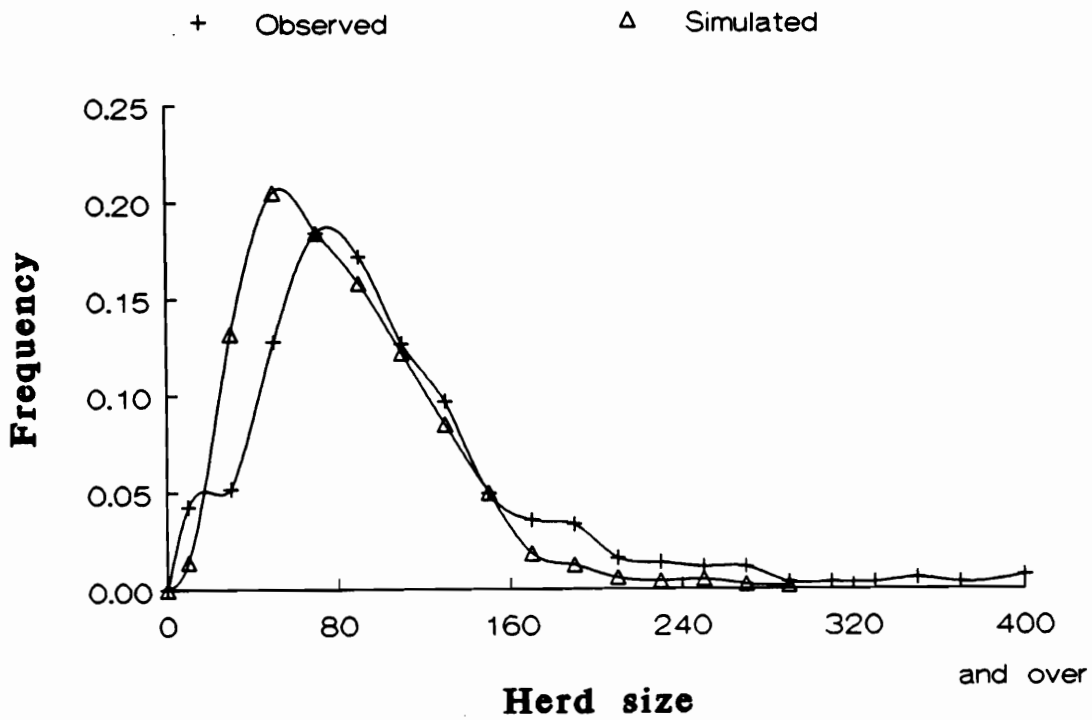
Figure 1. Distributions of herd sizes from Virginia DHI herds and from simulated populations.
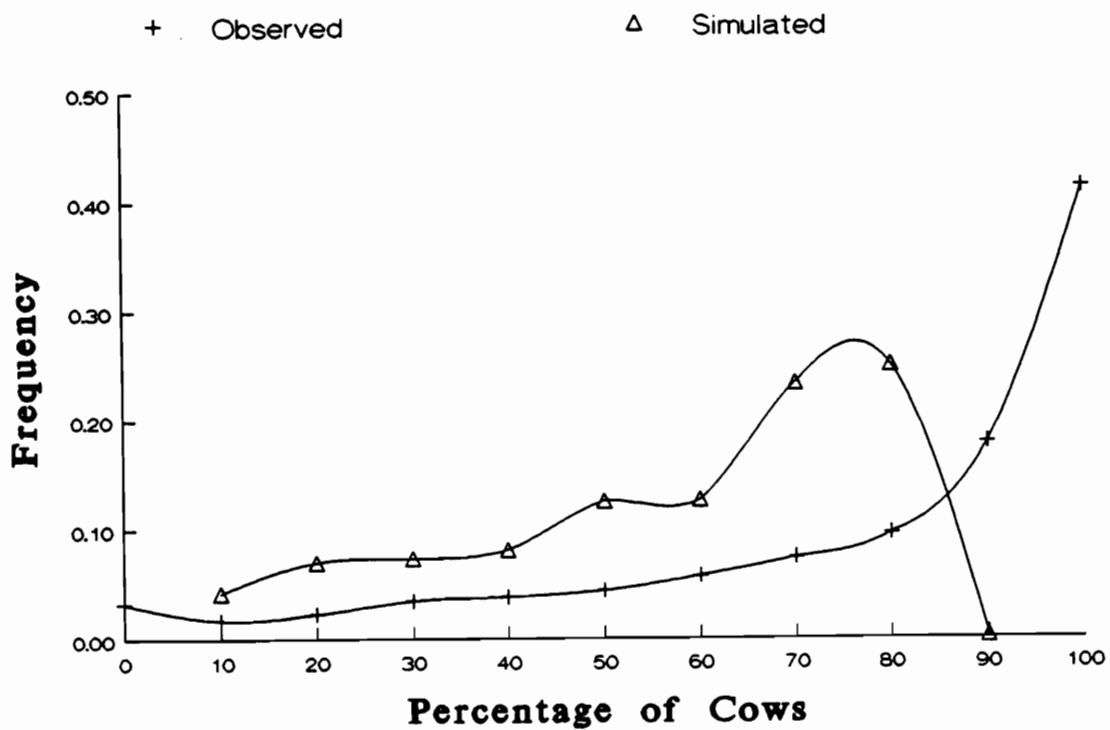
Figure 2. Distributions of proportions of sire-identified cows for Virginia DHI herds and from the simulated "medium" level of sire identification.

uniform random number for each cow. The numbers of animals in simulations were limited by computing time for genetic evaluations which increases geometrically as the number of animals increases arithmetically. Numbers of animals were controlled by varying the number of herds in each simulated population where herd sizes were determined by random numbers generated from an actual distribution. Numbers of sires were determined to approximate the average number of daughters for young sires in a initial proof. Sections a) through d) in Table 1 correspond to different study objectives, described subsequently. In section a), b), and c), sixty herds and 250 sires were simulated in each population with random herd sizes and proportions of sire-identified cows. Accordingly, numbers of identified daughters per sire differed by levels of sire identification. In section d) various numbers of herds and sires were simulated for populations with assigned herd size and proportion of sire identified cows so as to yield similar numbers of cows and daughters per sire, as shown in Table 1.

Random Number Generation

Generation of random numbers is an essential part of any simulation study. The numbers generated are not truly random, since they result from a deterministic process. However, they have precise probabilistic occurrences and variability according to the adopted distribution function for random numbers. The International Mathematical and Statistical Language (IMSL) collection of subroutines provides several random number generation subroutines for Fortran programs. These subroutines require an initial value to activate number generation. Initial values may be selected by using either a subroutine or the system clock on the mainframe. The mainframe system clock was used in this study. The random numbers used and their distributions were as follows.

Model effects on milk yield	normal distribution
Selection of animals for random mating	uniform distribution
Chance of first lactation from a mating	uniform distribution
Chance of culling after each lactation	uniform distribution
Deletion of sire identity of cows	uniform distribution
Determination of herd size	general distribution
Determination of proportion of cows with sire identity per herd	general distribution

A general distribution is an actual distribution from observed data and generation of random numbers from a general distribution will be explained subsequently. Effects such as those in the model describing milk yield are usually assumed to be normally distributed. Normal random deviates are required to simulate such effects. The normal random number generator, RNNOF, in IMSL (29) was used to produce standard normal deviates with mean 0 and variance 1.

The IMSL uniform random number generator, RNUN (29), produces numbers from 0 to 1. Within the interval 0 to 1, the probability that a random number falls in any subinterval is the same for all subintervals of the same length. The characteristics of the uniform distribution were utilized to make decisions with a certain probability of occurrence. For random mating, the probability that a given sire was chosen for mating to a given cow was equal for all sires and thus random numbers from equal subintervals were used to choose a sire identification. Other probabilistic decisions such as culling (or death) of cows after each lactation or giving birth to a female offspring which completed her first lactation, were made by comparing uniform random numbers with subintervals corresponding to the desired probabilities.

Simulated distributions should follow observed data to the extent possible. Figure 1 shows the distribution of herd size of Virginia DHI herds in 1990 and the distribution of simulated herd sizes. The distributions are skewed toward small herd size from a normal distribution. The actual distribution was used to produce randomly generated

herd sizes, as an alternative to fixed, assigned sizes, for parts of this study. An interpolation technique (1) was used with sets of cumulative frequency points from the actual population to produce a continuous distribution through the given points. Random numbers from a general continuous cumulative distribution were obtained from subroutines RNGCS and RNGCT in IMSL (29). Figure 2 shows actual and simulated distributions of proportion of cows in a herd with sire identification for Virginia DHI herds in 1990. The cumulative distribution of actual proportions was manipulated to yield simulated cumulative frequency points of not less than .1 or greater than .9 for populations with the "medium" level of sire identification.

Methods for Utilizing Records Lacking Sire Identity

Increasing accuracy of genetic evaluation in an animal model at reasonable cost is an important goal for animal breeders. Given the distribution of actual herd size (Figure 1), lack of mates in management group subclasses may seriously reduce the accuracy of estimated effects of management groups. Increasing the number of records within management groups may be advantageous in two ways. First, less combining of groups to achieve some required minimum group size could lead to groups more homogeneous in true group effect. Second, larger group size would reduce standard errors of estimated group effects and, consequently, prediction error variance of genetic evaluations.

Genetic groups (76, 92) have been used in the animal model to estimate genetic effects for offspring of animals, which lack certain information desired for genetic evaluation. In this study, several methods were considered to incorporate records of cows lacking sire identification and typically abandoned, treating their genetic effects in different ways. If additional records are included in management groups, less bias

due to required combining of groups and smaller prediction error variances due to increased number of records in groups are expected. Consequently, mean square errors of prediction (squared bias plus prediction error variance) may be reduced. The principal objective of this study was to test the use of records lacking sire information as group mates. At the same time, however, the influence on estimation of breeding values from records lacking sire identification should be carefully examined.

Four alternative methods were proposed to incorporate records lacking sire information into animal model genetic evaluations. Suppose the following records to illustrate the conventional and alternative methods

Group	Cow	Record
β_1	1	y_1
	2	y_2
	3 (no sire ID)	y_3

For the above example, the following model is given for simplicity.

$$y = X\beta + Za + e \tag{2.4}$$

where β and a are vectors of management group and additive genetic effects, and X and Z are incidence matrices, respectively. The record of cow 3 is excluded in the conventional method and the mixed model equations are as follows

$$
\begin{aligned}
2\beta_1 + a_1 + a_2 &= y_1 + y_2 \\
\beta_1 + (1 + \alpha)a_1 &= y_1 \\
\beta_1 + (1 + \alpha)a_2 &= y_2
\end{aligned} \tag{2.5}
$$

where α is the ratio of error variance to additive genetic variance and the a's are genetic values of cows. In the conventional analysis y_3 is discarded, even though the group size is thereby decreased, possibly requiring combining of groups. The mixed model equations for animals 1 and 2 after absorbing the group equation are

$$(.5 + \alpha)a_1 - .5a_2 = y_1 - (y_1 + y_2)/2$$
$$-.5a_1 + (.5 + \alpha)a_2 = y_2 - (y_1 + y_2)/2$$

The record, y_3, is not used in the adjustment for group effects and the daughter equivalents (86) of animals 1 and 2 are .5.

Method 1

The assumption proposed in method 1 to include the record of cow 3, is that the genetic value of cow 3 is population average. Population average may represent the average of genetic groups defined flexibly by birth date of animals to exclude genetic trend. In this study average genetic value of animals born in preceding 5 years was chosen to account for genetic values of records lacking sire identification. The mixed model equations in method 1 are as follows

$$3\beta_1 + a_1 + a_2 = y_1 + y_2 + y_3$$
$$\beta_1 + (1 + \alpha)a_1 = y_1 \qquad\qquad\qquad [2.6]$$
$$\beta_1 + (1 + \alpha)a_2 = y_2$$

In the first equation, the record of cow 3 is included and treated as the record of a group mate. There are equations for cows 1 and 2, but not for 3, since the assumption regarding the genetic value of the record does not require an equation for cow 3. The normal equations after absorbing the group effect are

$$(.67 + \alpha)a_1 - .33a_2 = y_1 - (y_1 + y_2 + y_3)/3$$
$$-.33a_1 + (.67 + \alpha)a_2 = y_1 - (y_1 + y_2 + y_3)/3$$

The daughter equivalents of animals 1 and 2 are increased from .5 to .67. In these equations, the record of animal 3 contributes to the genetic evaluations of animals 1 and 2, because the estimated group effect is a function of y_3. If a_3, the genetic value of the discarded animal, is not population average, then estimates of a_1 and a_2 are biased. In both equations, the term, $-.33a_3$, is omitted due to the assumption that $a_3 = 0$ relative to population average. A strong reason supporting the assumption of method 1 is that the expected value of the genetic effect of an animal with unknown pedigree is the average genetic value of the population.

Method 2

An alternative assumption is that the genetic value of cow 3 is the genetic average of the management group in which she makes her record such that $a_3 = .5(a_1 + a_2)$. The mixed model equations manipulated for method 2 are as follows

$$3\beta_1 + 1.5a_1 + 1.5a_2 = y_1 + y_2 + y_3$$
$$1.5\beta_1 + (1.5 + \alpha)a_1 = 1.5y_1 \qquad [2.7]$$
$$1.5\beta_1 + (1.5 + \alpha)a_2 = 1.5y_2$$

Equations [2.8] and [2.9] show the manipulation of mixed model equations. The absence of an equation for cow 3 results from the fact that the genetic value of cow 3 is a simple function of the genetic values of cows 1 and 2. Weighted coefficients from including cow 3 in the group total are given to both breeding values and records of

cows 1 and 2 without including the equation for animal 3. The expression of the weighted coefficients in matrix form is as follows

$$Z_w = Z + W$$

$$= \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} .5 & 0 \\ 0 & .5 \\ 0 & 0 \end{bmatrix} \qquad [2.8]$$

Columns and rows of Z_w represent cows with sire identification and records (with sire identification in the first and second rows and without in the third row), respectively, and the mixed linear model equations are

$$\begin{bmatrix} X'X & X'Z_w \\ Z'_wX & Z'_wZ + I\alpha \end{bmatrix} \begin{bmatrix} \beta \\ a \end{bmatrix} = \begin{bmatrix} X'y \\ Z'_wy \end{bmatrix} \qquad [2.9]$$

In the example, $X'y$ contains the record of animal 3 in addition to the records of animals 1 and 2. But Z'_wy includes only the records of animals 1 and 2 weighted by the increased numbers of group mates. The normal equations for animals 1 and 2 after absorbing are

$$(.75 + \alpha)a_1 - .75a_2 = 1.5y_1 - 1.5(y_1 + y_2 + y_3)/3$$
$$-.75a_1 + (.75 + \alpha)a_2 = 1.5y_1 - 1.5(y_1 + y_2 + y_3)/3$$

If the both right and left hand sides of above equations are divided by 1.5, the equations are

$$(.5 + 2/3\alpha)a_1 - .5a_2 = y_1 - (y_1 + y_2 + y_3)/3$$
$$-.5a_1 + (.5 + 2/3\alpha)a_2 = y_1 - (y_1 + y_2 + y_3)/3 \qquad [2.10]$$

The records of cows 1 and 2 are adjusted for the group effect estimated from all records including the record of animal 3. The fraction of variance ratio for cow 3 is eliminated, as illustrated by the term, $2/3\alpha$. In the sense of daughter equivalents, method 2 gives no increase relative to the conventional method.

Method 3

A third alternative is to account for the genetic value and permanent environmental effect of cow 3 through her producing ability rather than assume population or management group genetic average while ignoring permanent environmental effects. In method 3 producing ability is treated as a random effect. The model including producing ability for illustration of method 3 is

$$y = X\beta + Za + Zp + K\pi + e \qquad\qquad [2.11]$$

where a and p are vectors of additive genetic and permanent environmental effects of cows with sire identification, π is a vector of producing abilities of cows lacking sire identification, Z and K are incidence matrices corresponding to a, p and π. For records with sire identification, a one is assigned to the corresponding row of Z and all zero elements are assigned to the corresponding row of K, with the opposite occurring for records without sire identification. Each column of Z represents a cow with sire identification, while each column of K represents a cow without sire identification. The sum of permanent environmental (including non-additive genetic) and additive genetic effects is producing ability. For cows with sire identification, additive genetic and permanent environmental effects are estimated separately. For cows lacking sire identity, producing ability for the cow is included in the model. The mixed model equations for the example are as follows

$$3\beta_1 + a_1 + a_2 + p_1 + p_2 + \pi_3 = y_1 + y_2 + y_3$$
$$\beta_1 + (1 + \alpha)a_1 + p_1 = y_1$$
$$\beta_1 + (1 + \alpha)a_2 + p_2 = y_2$$
$$\beta_1 + a_1 + (1 + \alpha_p)p_1 = y_1 \qquad\qquad [2.12]$$
$$\beta_1 + a_2 + (1 + \alpha_p)p_2 = y_2$$
$$\beta_1 + (1 + \alpha_\pi)\pi_3 = y_3$$

where α_p is the ratio of error variance to permanent environmental variance, and α_π is the ratio of error variance to the sum of additive genetic and permanent environmental (including non-additive genetic) variances. The mixed model equations including records of cows lacking sire identity are written with matrix notation as

$$
\begin{bmatrix}
X'X & X'Z & X'Z & X'K \\
Z'X & Z'Z + A^{-1}\alpha & Z'Z & 0 \\
Z'X & Z'Z & Z'Z + I\alpha_p & 0 \\
K'X & 0 & 0 & K'K + I\alpha_\pi
\end{bmatrix}
\begin{bmatrix}
\beta \\ a \\ p \\ \pi
\end{bmatrix}
=
\begin{bmatrix}
X'y \\ Z'y \\ Z'y \\ K'y
\end{bmatrix}
\qquad [2.13]
$$

Relationships among cows lacking sire information, and among cows with and without sire identification are completely ignored, since the matrix $K'K + I\alpha_\pi$ is diagonal and off-diagonal block matrices between the cows with and without sire identity are null. In other words, while records of cows lacking sire information are used for estimating management group effects, producing ability of such cows is estimated without considering genetic relationships. Reasons for not considering genetic relationships in estimating producing ability include 1) the relatively large effect of p in π, and 2) the primary interest in improving estimates of group effect. Thus, producing abilities of cows lacking sire identity are connected only through management groups. As the number of lactations increases, more accurate estimation of producing abilities is ob-

tained, which implies improved accounting for genetic effects of cows lacking sire identification.

Method 4

In method 4 the producing ability of cow 3 is treated as a fixed effect, and the mixed model equations are as follows

$$
\begin{aligned}
3\beta_1 + a_1 + a_2 + p_1 + p_2 + \pi_3 &= y_1 + y_2 + y_3 \\
\beta_1 + (1 + \alpha)a_1 + p_1 &= y_1 \\
\beta_1 + (1 + \alpha)a_2 + p_2 &= y_2 \\
\beta_1 + a_1 + (1 + \alpha_p)p_1 &= y_1 \\
\beta_1 + a_2 + (1 + \alpha_p)p_2 &= y_2 \\
\beta_1 + \pi_3 &= y_3
\end{aligned}
\qquad [2.14]
$$

In the equation for animal 3, the variance ratio for producing abilities, α_π, is not considered, since fixed effects are assumed not to have a distribution and variance. The record of cow 3 is used in estimating group effects in methods 3 and 4 while producing ability is estimated. Estimated producing ability from method 3 may be used as a culling guide rather than an estimate of breeding value. However, similar use of estimates from method 4 would be appropriate only when comparing errors with equal numbers of records since producing ability is estimated as a fixed effect in method 4.

Genetic Evaluation

Genetic Evaluation System

An iterative procedure for genetic evaluation with an animal model was adopted for sparse linear systems to reduce computational requirements. The source code of the evaluation program is shown in the Appendix. The procedure used is referred to as the conjugate gradient method for a sparse linear system (71). Since iterative procedures do not involve inversion of the coefficient matrix, prediction error variances were not available. Known dispersion parameters used (heritability = .25 and repeatability = .50) in the simulation of records were used also in genetic evaluations. As in the USDA animal model (97), management groups were required in many parts of this study to have a minimum of 5 records, achieved where necessary by combining adjacent 2 month seasonal groups. Genetic evaluations with minimum group sizes of 3, 6 and 9 also were examined in part of the study to determine optimum minimum group size. Accuracy of genetic evaluation was measured by correlations between true and estimated breeding values, which are functions of mean square error of prediction. Correlations between true and estimated management group effects also were computed as measures of accuracy for management groups.

The model for the conventional genetic evaluation procedure is written with matrix notation as

$$y = X\beta + Za + Zp + e \qquad [2.15]$$

where X and β represent the incidence matrix and vector of management group effects. The incidence matrix, X, excludes records lacking sire identity, but these records are included in X of the alternative methods. The incidence matrices, Z, for the conven-

tional and alternative methods are not exactly equivalent since the Z's in alternative methods have rows of zeros for cows lacking sire identity, but both the conventional and alternative methods produce the same Z'Z. From equation [2.15], the mixed model equations are

$$
\begin{bmatrix}
X'X & X'Z & X'Z \\
Z'X & Z'Z + A^{-1}\alpha & Z'Z \\
Z'X & Z'Z & Z'Z + I\alpha_p
\end{bmatrix}
\begin{bmatrix}
\beta \\
a \\
p
\end{bmatrix}
=
\begin{bmatrix}
X'y \\
Z'y \\
Z'y
\end{bmatrix}
\qquad [2.16]
$$

Consequently, differences between the conventional and alternative methods were produced in X'X, Z'X, and X'Z. The Z'Z's in the mixed model equations for alternative methods are the same block matrices as the conventional method. Methods 3 and 4 additionally include producing abilities for cows lacking sire identification, and thus, the rank of the coefficient matrix in equation [2.13] is greater than that in equation [2.16].

Permanent environmental effects were absorbed into random additive genetic effects to reduce the size of the coefficient matrix, and, thereby, the number of rounds of iteration required for solution (71). Permanent environmental effects of sires were assumed to exist for convenience of absorption, even though no records are available for sires. Subtraction of the third row from the second in both the coefficient matrix and right hand sides of [2.16] gives

$$
\begin{bmatrix} 0 & A^{-1}\alpha & -I\alpha_p \end{bmatrix}
\begin{bmatrix}
\beta \\
a \\
p
\end{bmatrix}
= [0]
\qquad [2.17]
$$

For this algebra, the presence of permanent environmental effects for sires is required since the order of the identity matrix for variances of permanent environmental effects is equal to the order of the relationship matrix. After estimation of breeding values, permanent environmental effects are easily obtained as

$$\hat{p} = A^{-1}\hat{a} \times (\sigma^2_p/\sigma^2_a) \qquad [2.18]$$

As indicated by [2.18], since the inverse of the relationship matrix is available from the mixed model equations, permanent environmental effects are computed from the relationship matrix and the vector of breeding values. Permanent environmental effects for sires should be zero, because milk production records and non-additive genetic effects were not simulated for sires. However, estimates for sires were not exactly zeros due to rounding errors. Permanent environmental effects can be factored from [2.16] to reduce the number of iterations for solution (71) as follows using equation [2.18]

$$\begin{bmatrix} X'X & X'Z(I + A^{-1}\alpha_t) \\ Z'X & Z'Z(I + A^{-1}\alpha_t) + A^{-1}\alpha \end{bmatrix} \begin{bmatrix} \beta \\ a \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix} \qquad [2.19]$$

where $\alpha_t = \sigma^2_p/\sigma^2_a$. In method 3, producing abilities of cows lacking sire identification were included in the model as shown in equations [2.11] and [2.13]. The mixed model equations for method 3 after absorbing permanent environmental effects of records with sire identity from equation [2.13] were as follows

$$\begin{bmatrix} X'X & X'Z(I + A^{-1}\alpha_t) & X'K \\ Z'X & Z'Z(I + A^{-1}\alpha_t) + A^{-1}\alpha & 0 \\ K'X & 0 & K'K + I\alpha_\pi \end{bmatrix} \begin{bmatrix} \beta \\ a \\ \pi \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \\ K'y \end{bmatrix} \qquad [2.20]$$

After the absorption of π, the size of coefficient matrix was reduced as follows

$$\begin{bmatrix} X'X - X'K(K'K + I\alpha_\pi)^{-1}K'X & X'Z(I + A^{-1}\alpha_t) \\ Z'X & Z'Z(I + A^{-1}\alpha_t) + A^{-1}\alpha \end{bmatrix} \begin{bmatrix} \beta \\ a \end{bmatrix}$$
$$= \begin{bmatrix} X'y - X'K(K'K + I\alpha_\pi)^{-1}K'y \\ Z'y \end{bmatrix} \qquad [2.21]$$

However, even in the reduced coefficient matrix, the number of non-zero elements was dramatically increased relative to the conventional method or methods 1 and 2. This greatly increased the number of rounds of iteration required for solution (71). From the mixed model equations [2.21] the solution for estimated producing abilities for cows lacking sire identity is

$$\hat{\pi} = (K'K + I\alpha_\pi)^{-1}(K'y - K'X\hat{\beta}) \qquad [2.22]$$

Since the matrix $(K'K + I\alpha_\pi)$ is diagonal, its inverse is simply the diagonal matrix which has reciprocals of its elements on the diagonal. Producing abilities of cows lacking sire identification are easily estimated by the above equation. Method 4 does not require $I\alpha_\pi$ in equations [2.21] and [2.22], since producing ability is treated as a fixed effect.

Daughters and Sires in Genetic Evaluation

The potential to increase accuracy of estimating genetic merit through alternative evaluation methods is much greater for sires or cows with few offspring than for well proven sires having many daughters and other relatives. The objective of this study was therefore to simulate early evaluations of young bulls with limited progeny. The distribution of number of daughters per sire in simulations is shown in Figure 3 and Table 1. For the objective of directly comparing accuracies of the conventional and alterna-
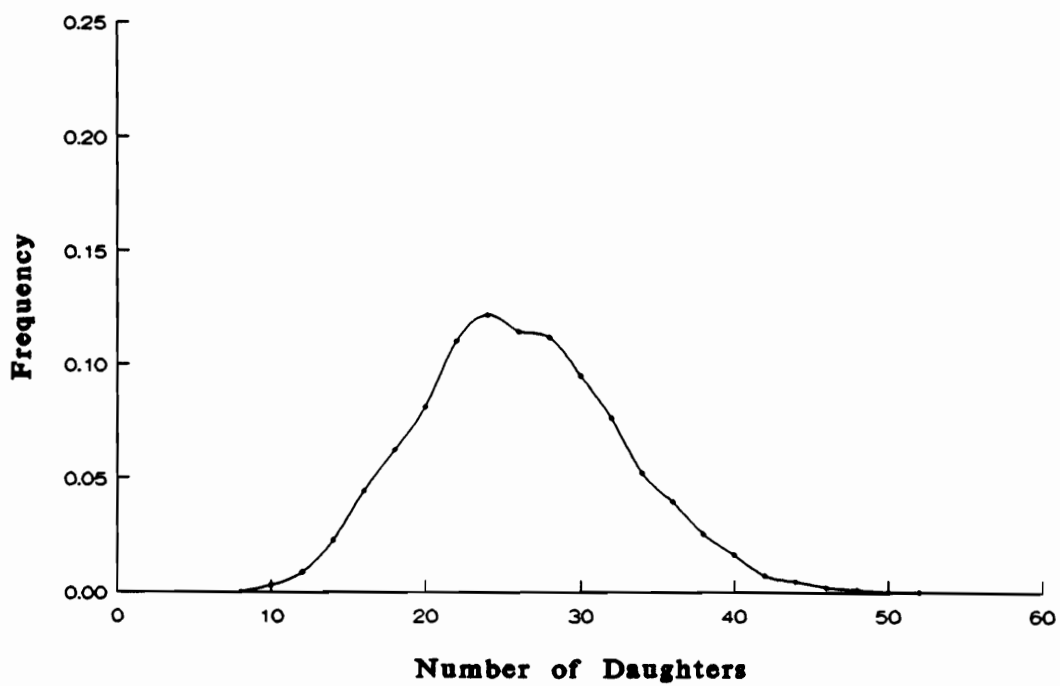
Figure 3. Distribution of numbers of daughters per sire in simulated populations.

tive methods, the mean number of daughters was 25.63 with standard deviation 6.56 and a minimum number of 9 (section b) in Table 1. Numbers of daughters per sire were similar for other objectives of the study. Young sires generally are not intensively selected at the initial proof, due to the low reliability of the first genetic evaluation. However, improved accuracy of early genetic evaluations for young sires could increase culling rate at the first proof and thereby reduce costs for maintaining or sampling young sires. Similarly, selection of cows as dams of young sires is based on limited information. Genetic evaluation methods which increase accuracy of estimated breeding value for potential dams of young sires may therefore increase the rate of genetic improvement.

Breeding Values and Permanent Environmental Effects of Sires

The distributions of true and estimated sire breeding values simulated for 9,000 sires are shown in Figure 4. The means of true and estimated breeding values were nearly identical, but standard deviation were 527.7 and 400.1kg for true and estimated breeding values, respectively. Reduced variation of estimated breeding values was expected since these were estimated as conditional means of true breeding values (i.e., the estimates are regressed toward the population mean).

Predictions of permanent environmental effects of sires are excluded from genetic evaluations for milk yield in dairy cattle, since records for milk yield are not observed for sires. However, the assumption of permanent environmental effects for sires provided simplicity in computation. If an animal had no record, the animals' permanent environment effect should be zero since non-additive genetic effects were not simulated. Equation [2.18] for obtaining permanent environmental effects after evaluation for additive genetic effects produced non-zero estimates of permanent environmental effects
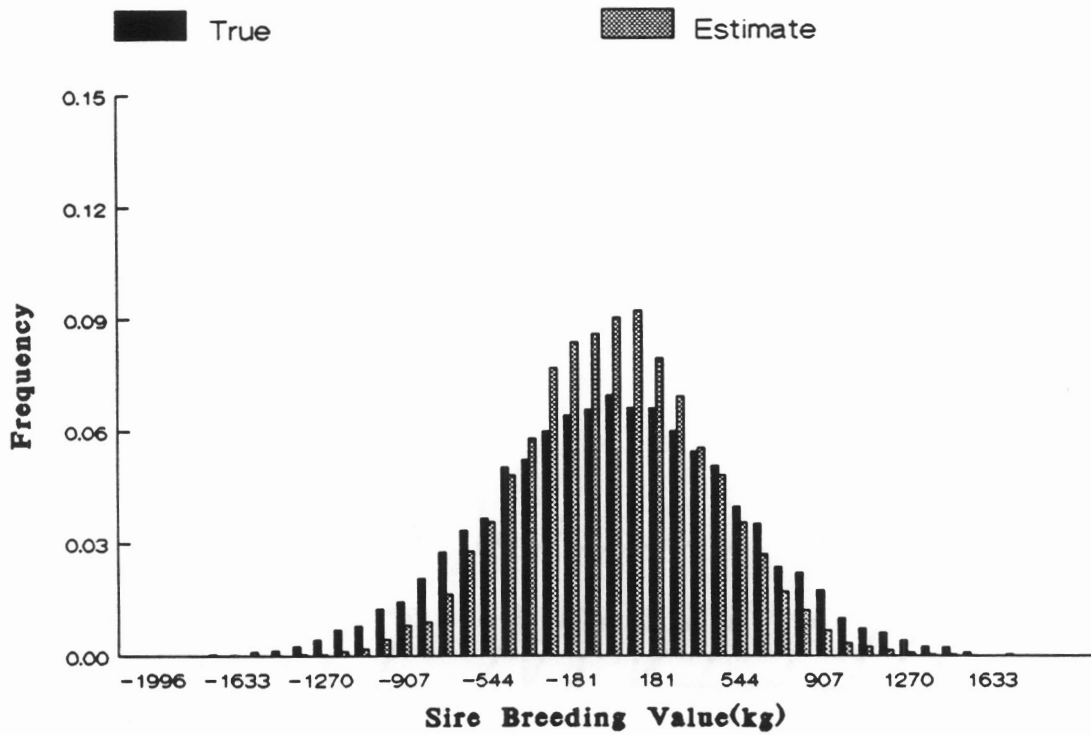
Figure 4. Distribution of true and estimated breeding values of sires.

for sires. Average permanent environmental effect was very close to zero (.098kg) with standard deviation of 3.54kg. Non-zero estimated values were considered as rounding errors in computation and are shown in Figure 5.

A Limit of the Conjugate Gradient Method for Sparse Linear Systems

The conjugate gradient method was used to solve the mixed linear model equations for genetic evaluations. The program for this algorithm was designed to avoid storing and working with zero elements in the coefficient matrix and, thus, to reduce computing storage and time. However, this method provides no guarantees regarding the number of iterations to converge or regarding accuracy of solution. The technique of singular value decomposition for a square matrix provided some indication of the number of iterations to convergence and accuracy of solutions. Suppose a mixed linear model as

$$Cb = r \tag{2.23}$$

By singular value decomposition for a square matrix the coefficient matrix C can be expressed as

$$C = U[\mathrm{diag}(w_i)]V' \tag{2.24}$$

where U and V′ are orthogonal matrices and $\mathrm{diag}(w_i)$ represents a diagonal matrix. Then the inverse of the coefficient matrix is

$$C^{-1} = V[\mathrm{diag}(1/w_i)]U' \tag{2.25}$$

The "condition-number" of a matrix is defined as the ratio of the largest w_i to the smallest w_i in equation [2.24]. A large condition-number indicates both an increase in
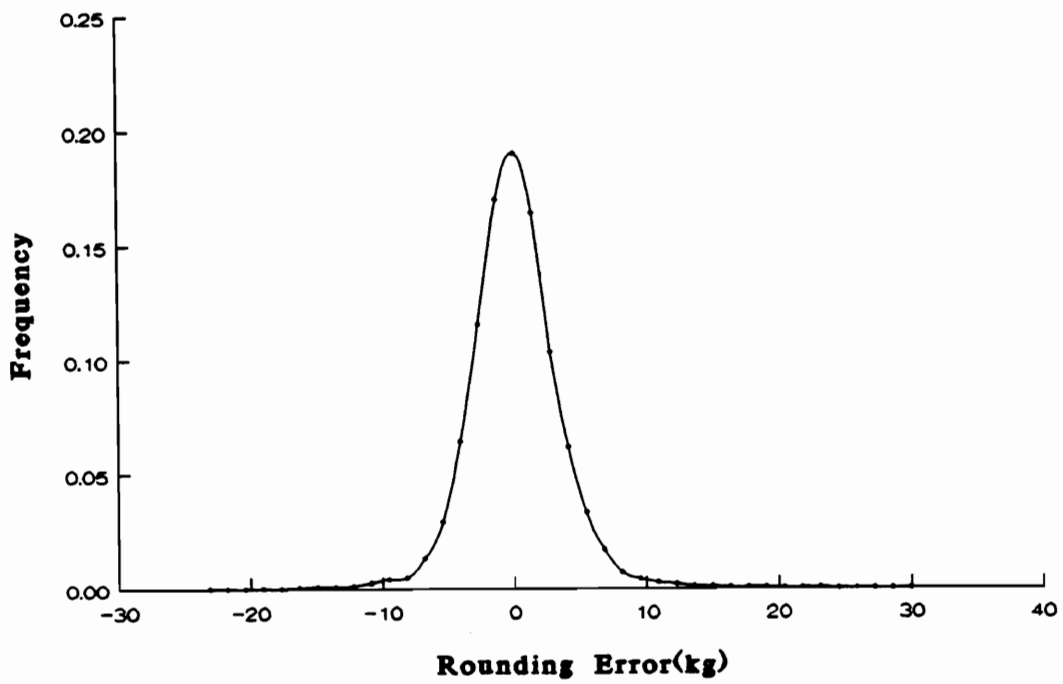
Figure 5. Rounding errors of sires' permanent environmental effects.

the number of iterations and reduced accuracy of solution. If C in equation [2.23] is symmetric and positive definite, a relatively smaller condition-number will be obtained (71). However, the coefficient matrices in equations [2.19] and [2.21] were not symmetric due to the absorption of permanent environmental effects. Further, in methods 3 and 4, producing abilities were absorbed into management group effects (equation [2.21]), yielding many non-zero elements (Table 2).

In genetic evaluations for 60 populations computed for the objective of directly comparing conventional and alternative methods of evaluation (Table 1, part b), there were 6 instances in which evaluations by methods 3 and 4 produced unreasonable solutions. A common feature of all was an extreme, negative estimate (e.g., -55,874.37) for the effect of the last management group in the last herd in genetic evaluations. In each case, the apparently erroneous solution was discarded, the evaluations for conventional and alternative methods recomputed, and "reasonable" estimates obtained. Table 2 compares correlations between true and estimated effects for apparently erroneous solutions (methods 3 and 4) and an unaffected solution set (method 1). Average number of non-zero elements in coefficient matrices was greater in methods 3 and 4 than method 1 due to absorbing producing abilities of cows lacking sire identifications. More non-zero elements in the sparse linear systems increased numbers of iterations for convergence (71) and, further, a larger condition-number due to the non-symmetrical coefficient matrices (71), possibly caused "unreasonable" solutions which consequently resulted in lower correlations between true and estimated management effects and breeding values for cows and sires. Genetic evaluations by method 3 seemed less affected than those by method 4, probably due to the variance ratio for producing abilities of cows lacking sire identification.

Table 2. Average number of non-zero elements and correlations between true and estimated manage-
ment group effects and breeding values in the failures of "reasonable" solutions for methods
3 and 4.

	Method 1	Method 3	Method 4
Number of non-zeros	46,617	103,212	103,212
	Correlations		
Management group	.756	.547	.403
Cow breeding value	.659	.587	.501
Sire breeding value	.836	.785	.711

Evaluation of Alternative Methods

Genetic evaluations by conventional and four alternative methods were used in computing correlations between true and estimated group effects and breeding values. Increases in correlations for management and genetic effects estimated by the alternative methods over those estimated by the conventional method would indicate improvements in accuracy obtained by the alternative methods. Four specific questions were addressed by analyzing correlations or gains in the correlation as described in the following four subsections. Characteristics of simulations for the four questions are presented in Table 1 (a through d).

1. Determination of Optimum, Minimum Group Size for Management Groups

The USDA implementation of the animal model "requires" a minimum of 5 records in management groups. Ideally, management groups should be as homogeneous as possible with respect to all effects except additive genetic and permanent environmental effects, and as large as possible with respect to number of records. Since these ideals are conflicting in practice, some compromise is required. Accuracy of estimating management group effects and breeding values by the alternative methods with regards to minimum group size was examined. Larger numbers of records in management groups for alternative methods relative to the conventional method should decrease prediction error variance and requires less combining of groups. However, potentially erroneous assumptions regarding genetic abilities of cows with unknown sires may offset these advantages. Accuracy was measured by correlations between true and estimated management group effects and breeding values, which are functions of mean square error (sum of prediction error variance and square of bias). Minimum group sizes of

3, 6, and 9 were examined. The model for correlations and increases in correlations (alternative minus conventional methods) was as follows

$$y_{ijk} = p_i + r_{ij} + s_k + e_{ijk}$$

where y is the correlation or increase in correlation between true and estimated group effects and breeding values, p is the proportion of cows with sire identification, r is replication, and s is minimum group size. Three levels for proportion of sire-identified cows were designated low, medium and high. The medium level was generated from Virginia DHI data in 1990, shown in Figure 2. Each level had the same general distribution, but the mean proportion differed by $\pm$ .1. As shown in Table 1 (section a), ten populations (replications) of 60 herds were simulated for each level of sire identification and herd sizes were randomly distributed with the general distribution (average = 81.9) shown in Figure 1. Averages of 24,362 records and 11,366 cows were simulated. Sire identity was deleted from cows to yield three levels of sire identification which averaged 57.1% across the three levels. A total of 250 sires with approximately 20 to 30 daughters each were simulated. Genetic evaluations were conducted by conventional and alternative methods for each population with minimum group sizes of 3, 6 and 9.

2. Improvement of Accuracy by Alternative Methods

The accuracies of conventional and alternative methods for estimation of group effects, cow breeding values, and sire breeding values were measured by correlations between true (simulated) and estimated effects. Estimated effects were from animal model genetic evaluation as described previously, with minimum group size of five. Analogous to simulations in the previous section, three levels of sire identity from the

general distribution in Figure 2 were simulated to provide variation in proportion of sire-identified cows. A total of 250 sires with approximately 20 to 30 daughters each were simulated. Sixty populations with averages of 23,849 records and 11,349 cows were simulated and deletion of sire identification for three levels produced an average of 6,399 cows per population (57.4%) with sire identification. Twenty populations of 60 herds were simulated for each level of sire identification (Table 1, section b). Correlations between true and estimated effects were averaged for all 60 populations for conventional and four alternative methods, and differences between correlations by alternative and conventional methods computed. Means, ranges and distributions of differences between each of the four alternative methods and the conventional method were computed for management group effects and breeding values of cows and sires. Null hypothesis of no difference between each alternative method and conventional method were tested by t-tests.

3. Impact of Missing Identifications of Lower Breeding Value Sires

No published evidence exists to suggest that omissions of sire identification are other than examples of poor record-keeping which are randomly distributed with respect to sire breeding value. However, under certain situations, it may be advantageous to intentionally omit identification of sires with lower breeding values. To investigate the potential effects of intentional omission of identifications of sires with lower breeding values, three different percentages (0, 10 and 20%) of omitted lower breeding value sires were examined. The model for analysis of correlations or increases in correlations was as follows

$$y_{ij} = p_i + r_j + e_{ij}$$

where y is correlation or increase (alternative minus conventional) in correlation between true and estimated group effect and breeding values, and p is the proportion of intentionally missing lower true breeding sires from all sires.

Five populations (replications) were simulated and each population included 60 herds with an average of 24,917 records and 11,644 cows (Table 1, section c). Sire identification was randomly deleted as previously to yield, in the first phase, an initial population in which 65% of the cows (7,566 of 11,644) were sire identified. Three sub-populations were formed from the initial population by deleting identification for 50 sires chosen randomly, 25 chosen randomly and 25 lowest true breeding value sires, and 50 lowest true breeding value sires. This produced sub-populations in which identity was intentionally deleted for the lowest 0, 10, and 20% of sires based on true breeding values. Average numbers of sire-identified cows after the second phase of deletion were 6186, 6085 and 6045 for 0, 10 and 20% of intentional deletion of lowest breeding values, respectively (Table 1, section c). Figure 6 shows distributions of true breeding values of sires chosen for the second phase of deletion (i.e., breeding values of sires whose identification was deleted). With 20% intentional deletion, breeding values of sires with deleted identity averaged -715kg. With 10% and 0% intentional deletion, breeding values averaged -410kg and -11kg, respectively. Five populations were simulated for replication.

4. Performance of Alternative Methods in Individual Herds

In theory, any relative advantage of the alternative evaluation methods examined in this study should result from more accurate estimation of environmental effects of more homogeneous environmental groups, and the removal of these effects from genetic evaluations. It is reasonable to suppose that the potential to increase estimation
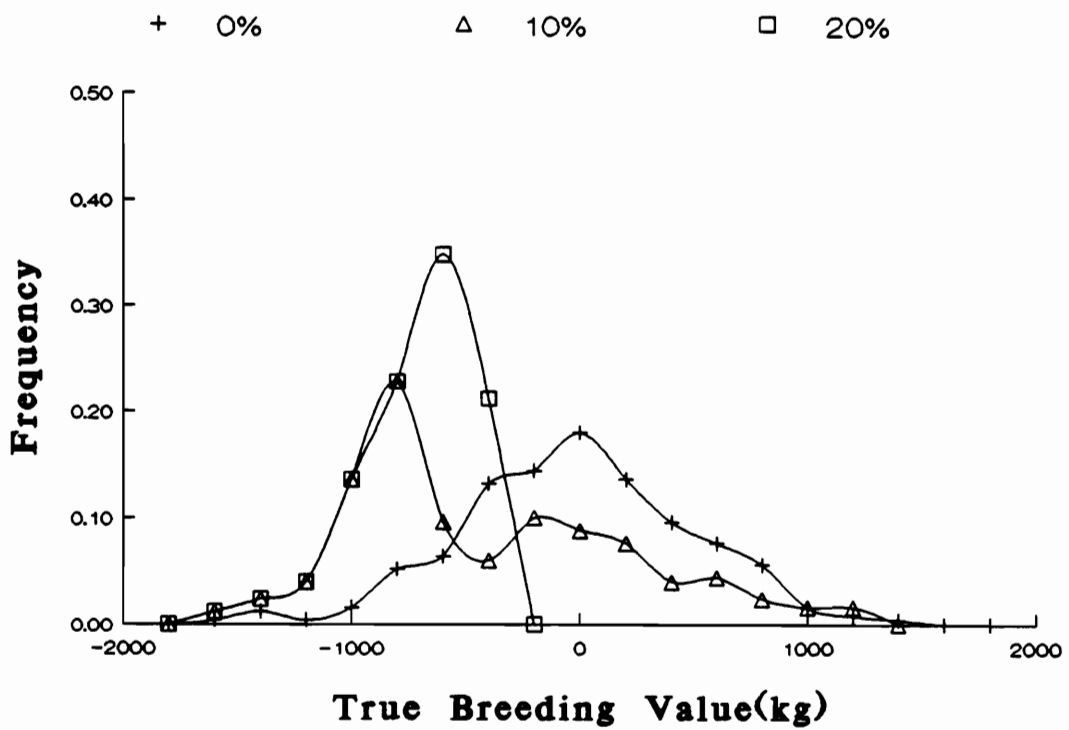
Figure 6. Distribution of true breeding values of intentionally and unintentionally selected sires.

accuracy and homogeneity of environmental effects will vary across herds such that it may be both beneficial and practical to employ different alternative methods in different herds, according to various herd characteristics.

Herd size, proportion of sire identified cows, and level of temporary herd effects are herd characteristics with major effects on numbers and homogeneity of records in subclasses. Herd size as shown in Figure 1 and proportions of sire identified cows as shown in Figure 2 vary considerably and, therefore, so also will numbers and homogeneity of records in management groups formed in these herds. Heterogeneous variances of environmental effects within herds (90) imply herd-to-herd differences in the relative importance of environmental factors and the consequences of combining environmental groups. Alternative methods which differ in their strategies to account for genetic values of cows lacking sire identity may therefore be more or less advantageous in a given herd according to identifiable characteristics of the herd.

Herd size is a major factor in determining the number of records in herd-year-season (2 month) subclasses. In large herds, numbers of records may be sufficient to estimate management group effects without combining groups, while relatively distant groups may need to be combined in small herds. In this context, three herd sizes of 40, 70 and 100 were examined for impact on accuracy of genetic evaluation from using records lacking sire identity. In contrast to previous simulations, herd sizes were fixed at 40, 70 and 100 and number of herds varied to yield similar number of cows. Populations with 80 herds (initial herd size = 40), with 60 herds (initial herd size = 70) and with 40 herds (initial herd size = 100) were simulated and averaged 15,602 records and 7,279 cows, 20,498 records and 9,628 cows, and 19,670 records and 9,175 cows, respectively (Table 1, section d).

Proportions of sire-identified cows in previous simulations varied from 0 to 100% (three levels of sire identified cows) across herds. To examine more specifically the ef-

fect of proportion of sire identification on performance of alternative methods, three fixed proportions (80, 50 and 20%) of sire-identified cows were simulated for each herd size. Numbers of sire identified daughters per sire were maintained at approximately 25 by simulating different numbers of sires according to the proportion of cows lacking sire identification and population size. A total of 230, 140 and 50 sires were simulated for 80, 50 and 20% sire-identified cows, respectively, at herd size 40, and 300, 180 and 70 sires, respectively, at herd sizes 70 and 100. The average numbers of daughters per sire were shown in section d) of Table 1.

Temporary herd effects include a variety of effects which are common to all cows calving in the same herd, year and season (2 month), but which vary across year seasons in the same herd and across herds. As shown previously, temporary herd effects contribute to the management group effect and were simulated to have a standard deviation of 522 kg. Greater variance in temporary herd effects will increase management group variance and thereby affect accuracy of genetic evaluation. The effect of increased variance in temporary herd effects on conventional and alternative methods including records lacking sire identification was therefore investigated by increasing the standard deviation of temporary herd effects to 733 kg.

The model for analysis was as follows

$$y_{ijkl} = h_i + p_j + v_k + r_l + hp_{ij} + hv_{ik} + pv_{jk} + hpv_{ijk} + e_{ijkl}$$

where y is the increase in correlation between true and estimated group effects and breeding values by alternative methods, h represents herd size, p represents proportion of sire identification, v represents level of variance in temporary herd effect, r is replication ($l = 1,2,3$), and terms described by two or more letters in the model represent interactions.

For this part of the study, simulation parameters were determined by first setting herd sizes and proportions of sire-identified cows of interest. Next, the numbers of herds were determined to produce similar numbers of cows in populations. Numbers of sires were decided to yield approximately 25 identified daughters per sire. Six replications (populations) of each herd size (40, 70 and 100) and sire ID level (20, 50 and 80%) shown at section d) in Table 1, included three populations simulated with low variance $(552kg)^2$ and another three populations with high variance $(733kg)^2$ of temporary herd effects. The desired proportions of sire-identified cows were obtained by deleting sire identity using random numbers.

RESULTS AND DISCUSSION

Four alternative methods of using records lacking sire identification were compared with conventional animal model genetic evaluation with regard to four major questions. These questions included determination of an optimum minimum management group size to obtain maximum evaluation accuracy, the probability and distribution of increased accuracy by using alternative methods, the impact on accuracy of intentionally omitting identifications of lower breeding value sires and the performance of alternative methods under varying conditions of herd size, proportion of sire identified cows, and variance of temporary herd effects. Results for each question are presented sequentially.

1. Determination of Optimum, Minimum Size for Management Groups

Management groups are defined in genetic evaluation to remove effects of environmental factors common to all individuals within a group but which vary across groups. To minimize within group variation, preference is given to groups which are relatively restricted in time, location, or other dimensions corresponding to variation in environmental conditions. Establishing a minimum number of records for management groups represents a somewhat arbitrary trade-off between desired within group homogeneity and desired increased accuracy of effect estimation from larger numbers of records per group. Optimum minimum group size may vary according to herd sizes and proportions of sire-identified cows for each herd, since these are major factors determining management group size and the frequency of required combining of groups.

Number of records in management groups and frequency of combining groups influence accuracy of genetic evaluation because accuracy of genetic evaluation partly depends on how well estimated management group effects remove group environmental conditions from genetic evaluations. In the current simulations, factors used to define management groups were herd, year, and 2 month-season, which produced six possible management groups in each herd-year. From 30 simulated populations with various herd sizes and three levels of sire-identified cows, genetic evaluations were computed by conventional and alternative animal model evaluation methods with minimum group sizes of 3, 6 and 9 as shown in a) of Table 1.

Table 3 shows numbers of herd-year-season and management groups and average numbers of records per group according to minimum group sizes. Alternative methods obtained additional 79 herd-year-season groups which lacked records with sire identification. Alternative methods required less combining of adjacent groups and provided more records in management groups than the conventional method for all levels of minimum group size. This was expected since alternative methods used records lacking sire identity not used by the conventional method. Differences in numbers of management groups between alternative methods 1 and 2 vs. 3 and 4 were from eliminating management groups lacking any records with sire identity in methods 1 and 2. Management groups of alternative methods 3 and 4 included only 3, 12, and 28% of records from adjacent 2-month seasons with minimum group sizes of 3, 6 and 9, respectively, while management groups of the conventional method contained 13, 36 and 71% from required combining of groups to achieve minimum group size. The alternative methods also yielded approximately 13, 28, and 44% more management groups and 50, 34, and 21% more records per management group with minimum group sizes of 3, 6, and 9, respectively.

Table 3. Number of management groups and average number of records in groups for conventional and alternative methods according to minimum group sizes in simulated populations.

Minimum group size		Conventional method		Methods 1 & 2		Methods 3 & 4	
	H-Y-S groups	1709		1788		1788	
	Records per group	8.4		13.6		13.6	
3	Management groups	1512		1680		1741	
	Records per group	9.5	(13%)[a]	14.2	(4%)	14.0	(3%)
6	Management groups	1211		1550		1592	
	Records per group	11.4	(36%)	15.4	(13%)	15.2	(12%)
9	Management groups	964		1374		1401	
	Records per group	14.4	(71%)	17.4	(28%)	17.4	(28%)

a Percent of records from other herd-year-2 month season groups.

Table 4. F-statistics for the effect of minimum group size and means of correlations between true and estimated group effects and breeding values by minimum group size.

	Group size	Conventional method	Method 1	Method 2	Method 3	Method 4
Group effect		215.37	81.94	81.94	35.56	11.38
	3	.7234	.7408	.7408	.7477	.7424
	6	.7313	.7429	.7429	.7489	.7437
	9	.7381	.7455	.7455	.7507	.7446
Cow breeding value		12.27	21.21	13.72	36.46	23.44
	3	.6401	.6488	.6436	.6514	.6486
	6	.6401	.6483	.6430	.6506	.6480
	9	.6385	.6472	.6422	.6492	.6464
Sire breeding value		.75[a]	3.53	2.27[a]	4.14	2.99[a]
	3	.8212	.8280	.8257	.8288	.8269
	6	.8206	.8275	.8252	.8283	.8266
	9	.8200	.8265	.8243	.8271	.8254

a Non-significant; others ($P < .05$)

Table 4 shows F-statistics for the effect of minimum group size and means of correlations between true and estimated group effects and breeding values by minimum group size. Slightly, but consistently higher correlations between true and estimated group effects and breeding values were obtained with alternative methods for all minimum group sizes. Correlations between true and estimated group effects tended to increase with increasing minimum group size in both conventional and alternative methods, even though the magnitudes of increase were very small. In contrast, correlations between true and estimated breeding values tended to decrease slightly with increasing minimum group size for both conventional and alternative methods. This implies that while larger management group size yielded more accurate estimation of group effects, homogeneity of groups was somewhat more important for accurate estimation of breeding values. F-statistics for correlations between true and estimated group effects and cow breeding values by all methods, including the conventional method showed significant differences due to minimum group size. This was not always true, however, for correlations between true and estimated breeding values of sires. Breeding values of sires estimated by the conventional method appeared not to be influenced by minimum group size, consistent with the finding of no difference in correlations between sons' daughter yield deviation and their parent evaluation with minimum group sizes of 5 or 10 (60). Larger F-statistics due to minimum group size were obtained for sire breeding values of methods 1 and 3, indicating significant differences ($P < .05$). As indicated above, accuracies of estimated sire breeding values were higher with smaller minimum group sizes for all methods, even though differences were very small. Differences in results for breeding values of sires and cows compared to group effects may result partially from the confounding of group size and group number. Specially, sire and cow breeding values may be slightly more accurate when estimated from a larger number of more homogeneous (and smaller sized) management

groups, while group effects, which span far fewer management groups, benefit more from larger group size.

Figure 7 shows increases in correlations (alternative minus conventional methods) between true and estimated group effects and breeding values according to minimum group size. Accuracy of estimated group effects was substantially increased by alternative methods especially with smaller minimum group size. Average number of records increased from 9.5 to 14.0 (50% increase) in management groups of minimum size 3 by alternative methods, while increases were approximately 30 and 20% with minimum group sizes of 6 and 9, respectively (Table 3). As mentioned previously, less need to combine records from different herd-year-seasons (3% in method 3) for minimum group size 3 yielded highest increases in correlations for management group effects. Increases in correlations between true and estimated cow and sire breeding values were without, however, apparent differences among minimum group sizes. F-statistics for effects of minimum group size on increases in correlations between true and estimated cow breeding values (Table 5) were greater than those for sire breeding values, but neither were significant ($P > .05$). Increases in accuracy of estimated management group effects may have contributed more to increased accuracy of cow breeding values by removing group environmental conditions from their records, since cows are less likely than sires to be tested over a range of environmental conditions.

2. Improvement of Accuracy by Alternative Methods

The accuracies of conventional and alternative methods for estimating management group effects, cow breeding values and sire breeding values were measured by correlations between true and estimated effects. Increases in correlations (alternative minus conventional methods) between true and estimated effects were considered as
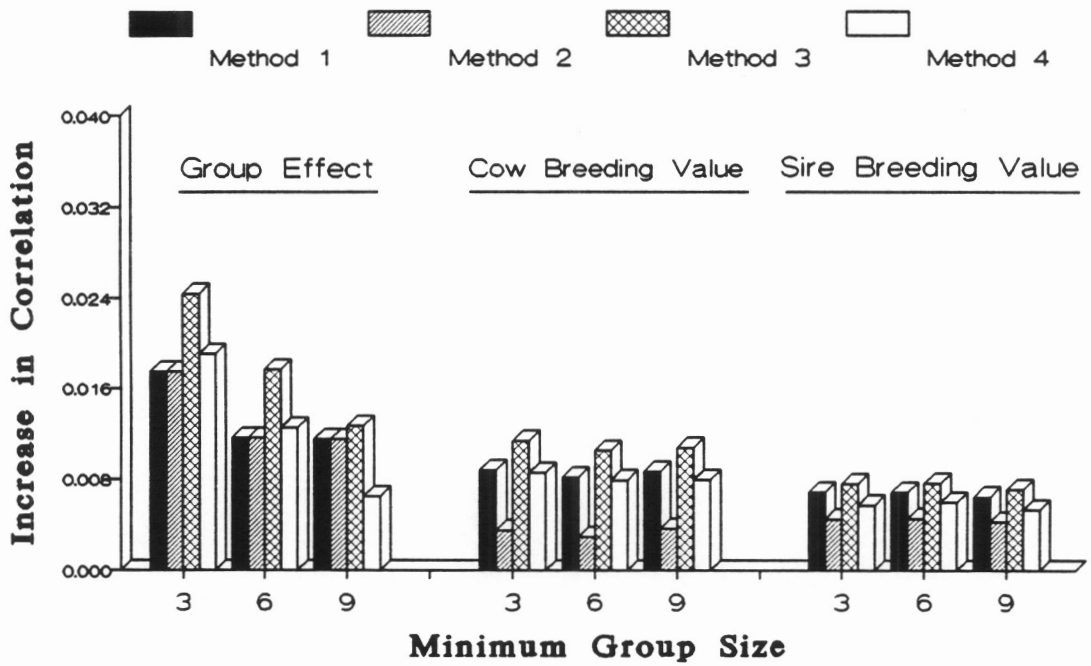
Figure 7. Increases, relative to the conventional method, in correlations between true and estimated group effects and breeding values according to minimum group size.

Table 5. F-statistics for the effect of minimum group size on increases of correlations between true and estimated group effects and breeding values due to minimum group sizes.

Source	Method 1	Method 2	Method 3	Method 4
Group effect	99.93	99.93	122.80	95.89
Cow breeding value[a]	1.34	1.79	2.22	1.21
Sire breeding value[a]	.16	.07	.26	.28

a Non-significant; others (P < .05).

improvements in accuracy by alternative methods. Size and number of simulated populations were limited by computing costs. Twenty populations, each with 60 herds, at low (47.4%), medium (57.7%) and high (67.9%) levels of sire identification were simulated to compare conventional and alternative methods of using records lacking sire identification. Differences between correlations using conventional and alternative procedures were computed for each of the total of 60 populations (across levels of sire identification) and an overall mean difference was computed for each of the four alternative methods. Twenty populations (replications) with three levels of proportion of sire identified cows, pseudo-random variation in herd size, and 250 sires for each population were simulated. Resulting averages were 23,849 records, 11,139 cows, 6,399 cows with sire identity (57.4%) and 25.63 identified daughters per sire as shown in b) of Table 1.

Table 6 shows the average number of herd-year-seasons and management groups and number of records per group in simulated populations. Numbers of herd-year-seasons represent those with one or more sire identified records for the conventional method and with one or more records, sire identified or not, for alternative methods. Comparisons between numbers of herd-year-seasons and management groups for conventional and alternative methods shows the relative loss of herd-year-seasons in forming management groups for genetic evaluation. On average for all populations, 84 herd-year-seasons (1783-1699) lacked sire identified records and therefore were lost in the conventional method, but were used by alternative methods. Also the numbers of records in herd-year-seasons increased from 8.1 to 13.4 which provided more observations, and thus, less combining of herd-year-seasons to obtain management groups of the required minimum size. In the conventional method, 398 herd-year-seasons (1699-1301) (potential management groups) were lost by required combining of adjacent herd-year-seasons, while only 137 (1783-1646) were lost in alternative methods 3

Table 6. Averge number of herd-year-seasons (HYS), records per HYS, management groups (MG) and records per MG according to levels of proportions of sire-identified cows in simulated populations.

Level[a]		Conventional method	Method 1 & 2	Method 3 & 4
Low	HYS	1618	1782	1782
	Records/HYS	6.9	13.2	13.2
	MG	1165	1534	1644
	Records/MG	9.5	14.5	14.3
Medium	HYS	1726	1783	1783
	Records/HYS	8.1	13.6	13.6
	MG	1314	1609	1638
	Records/MG	10.7	14.9	14.8
High	HYS	1756	1784	1784
	Records/HYS	9.3	13.5	13.5
	MG	1436	1648	1657
	Records/MG	11.4	14.5	14.5
Average	HYS	1699	1783	1783
	Records/HYS	8.1	13.4	13.4
	MG	1301	1596	1646
	Records/MG	10.5	14.7	14.5

a Level of sire identification. Averages are Low, 47.4; Medium, 57.7; High, 67.9%.

and 4. In methods 1 and 2, an additional 50 potential management groups were lost because no records with sire identification were in those potential management groups. In methods 3 and 4, even if there were no sire-identified records in a herd-year-season, a management group would still remain for estimating producing abilities of cows lacking sire identity as long as there was a minimum of 5 records. Methods 1 and 2 differ in that method 2 requires at least one sire identified record to determine genetic value for records not sire identified, while with method 1, there is no reason to estimate effects of management groups lacking sire identified records.

After herd-year-seasons were combined to form management groups, average numbers of records in management groups of the alternative methods were increased slightly, but more in the conventional method. The increase in average number of records in management groups through combining of groups represents a potential loss of homogeneity of the management group effect on records in the group. This potential loss was greatest in the conventional method. Table 6 also shows differences by level of sire identification in the necessity to combine herd-year-seasons to achieve minimum, required management group size. In the low level of sire identification, the difference between conventional and alternative methods in the necessity to combine herd-year-seasons (number of herd-year-seasons vs. number of management groups) was substantially greater than in the high level of sire identification.

Correlations between true and estimated management group effects and breeding values for cows and sires were used as measures of estimation accuracy. Differences in correlations between true effects and estimates computed by conventional and alternative methods indicate advantages or disadvantages for each alternative relative to conventional methods. As shown in Figures 8, 9 and 10, differences in correlations were small. However, relatively small but consistent differences in accuracy may have a meaningful effect in a large population over time.

Distributions of increases (alternative minus conventional methods) in correlations between true and estimated group effects from 60 populations for each method are shown in Figure 8. Only for method 4 were there any instances in which correlations by alternative methods were less than those by the conventional method (one of 60 populations). Differences in mean correlations for management groups (alternative minus conventional) ranged from .0187 for method 3 to .0119 for methods 1 and 2, and were significantly (P > .01) greater than zero by t-test for each of the four alternative methods.

Management group effects, additive genetic effects, and permanent environmental effects are estimated simultaneously in the mixed linear model. For the previous example (p 38), including permanent environmental effects for methods 1 and 2 here for illustration, management group solutions for the various alternative methods are

$$\hat{\beta} = (y_1 + y_2 + y_3 - \hat{a}_1 - \hat{a}_2 - \hat{p}_1 - \hat{p}_2)/3 \qquad \text{(Method 1)} \qquad [3.1]$$

$$\hat{\beta} = (y_1 + y_2 + y_3 - 1.5\hat{a}_1 - 1.5\hat{a}_2 - \hat{p}_1 - \hat{p}_2)/3 \qquad \text{(Method 2)} \qquad [3.2]$$

$$\hat{\beta} = (y_1 + y_2 + y_3 - \hat{a}_1 - \hat{a}_2 - \hat{p}_1 - \hat{p}_2 - \hat{\pi}_3)/3 \qquad \text{(Methods 3 and 4)} \qquad [3.3]$$

These equations are obtained from solving equations [2.6], [2.7],[2.12] and [2.14] for management group effects, $\hat{\beta}$. These results show two possible sources of bias in the estimation of management group effects for methods 1 and 2 which result from assumptions regarding the genetic value of animal 3 (lacking sire identification). If animal 3 has a different genetic value from the assumed population or group average, biased estimation of management group effects may result. Also the permanent environmental effect of animal 3, which is ignored (or assumed population average) in methods

Figure 8. Distribution of increases in correlations between true and estimated management group effects by four alternative methods from 60 populations.

1 and 2, may lead to biased estimation of the management group effect. Methods 3 and 4 may represent an improvement with respect to these problems, since the genetic effect and permanent environmental effect of animal 3 are considered through producing ability, π_3, in the estimation of the management group effect as shown in [3.3]. In general, alternative method 3, which considered producing abilities as random effects, gave the largest increases in accuracy of estimating management group effects.

Genetic evaluations of cows contribute more directly to the evaluation of sires in the USDA animal model genetic evaluation than in the previous genetic evaluation procedures. Accuracy of genetic evaluation of cows is limited, relative to sires, by fewer potential numbers of offspring. However, alternative methods may contribute more to cow evaluation through more accurate estimation of management group effects. Figure 9 shows distributions of increases in correlations between true and estimated cow breeding values by alternative methods. Increases in mean correlations for alternative methods ranged from .0101 (method 3) to .0020 (method 2) and were less than increases for management group effects, similar to that shown in Figure 7 of the previous section. Additionally, 18% (11 of 60 populations) of correlations by alternative method 2 were less than those for the conventional method of discarding records lacking sire identification. The assumption of method 2 that genetic values of cows lacking sire identification were management group average was the likely cause. Except for method 2, probabilities of negative increases in correlations were similar to those for group effects. Correlations from method 1 and 3 exceeded those from conventional method in all populations and in 59 of 60 populations with method 4.

Developing methods which potentially could increase accuracy of breeding values for young sires was an important objective of this study. Increased accuracy of genetic evaluation from progeny test results for young sires could enhance the rate of genetic progress and reduce costs of maintaining bulls in waiting. Increases in correlation be-
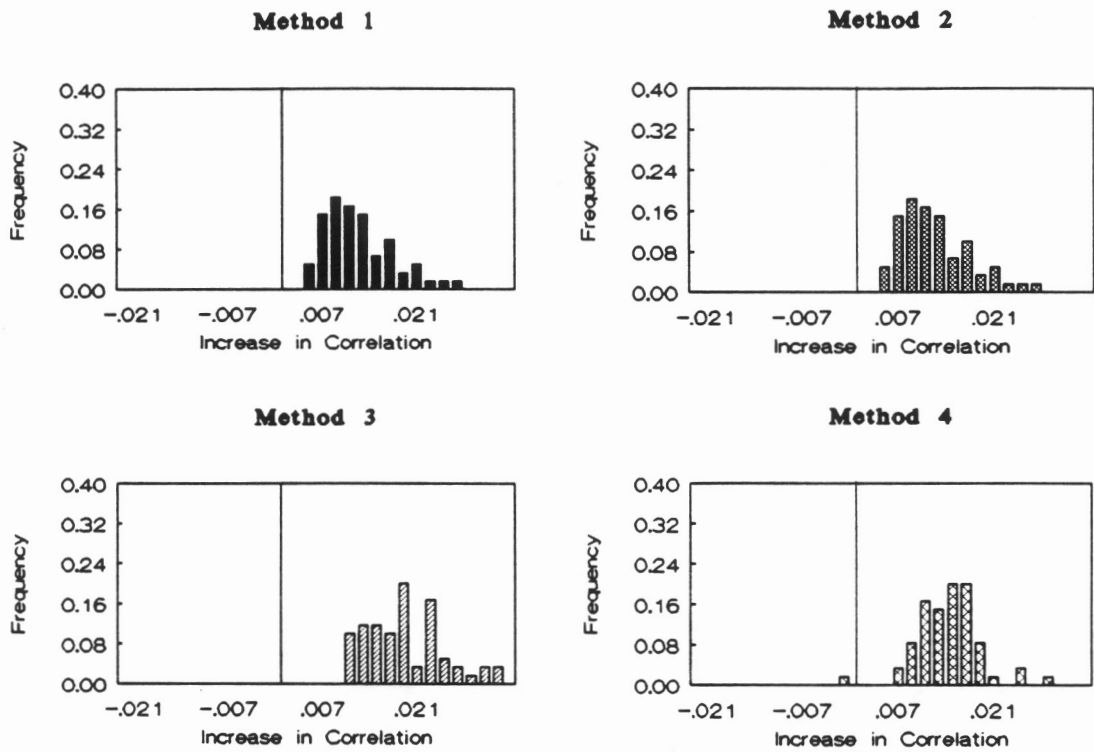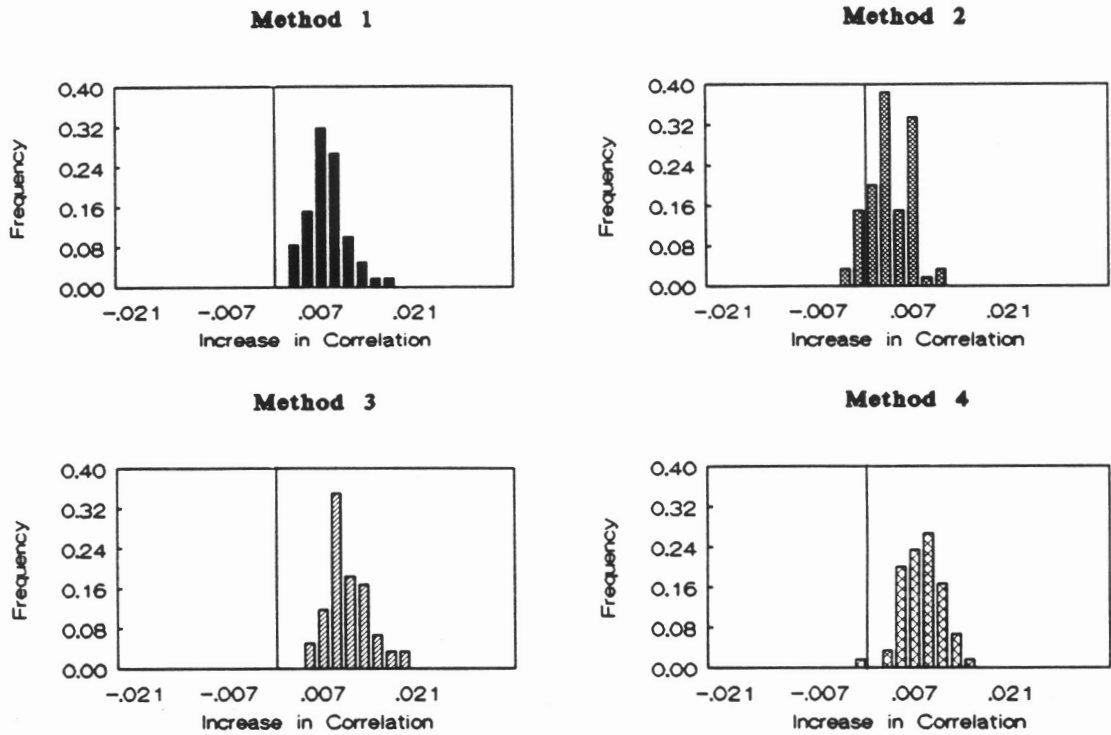
Figure 9. Distribution of increases in correlations between true and estimated cow breeding values by four alternative methods from 60 populations.
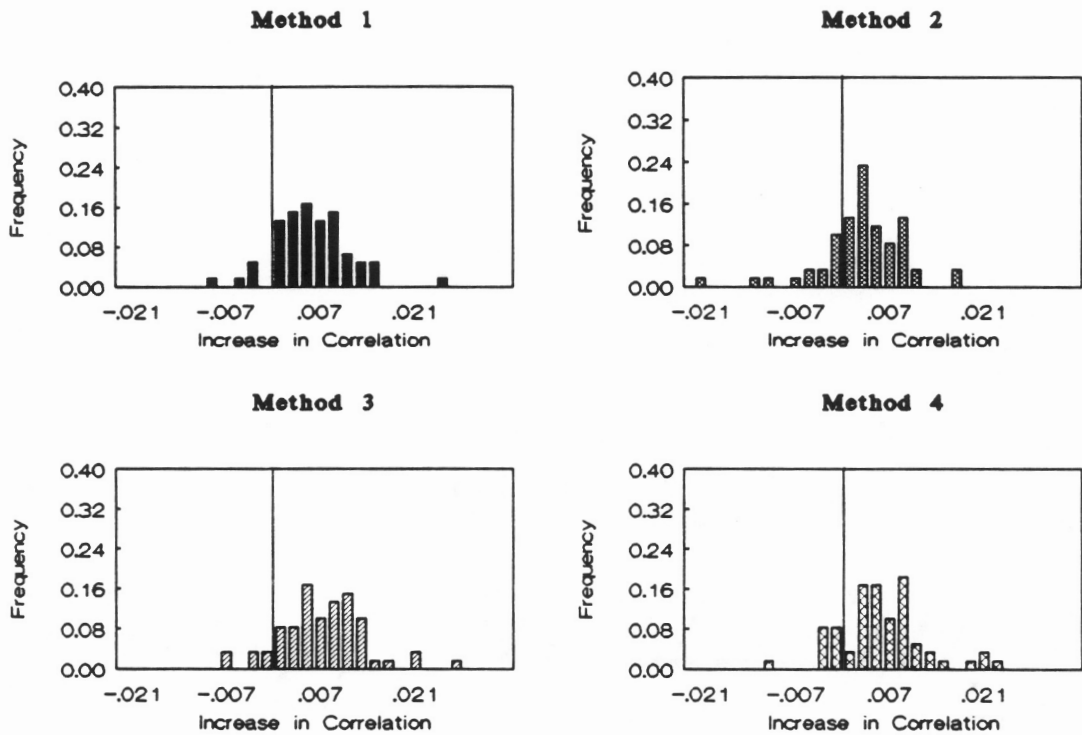
Figure 10. Distribution of increases in correlations between true and estimated sire breeding values by four alternative methods from 60 populations.

tween true and estimated sire breeding values were smaller and more variable than increases for cow breeding values, as shown in Figure 10. Increases in mean correlations by methods 1, 2, 3 and 4 were .0054, .0024, .0069 and .0053, respectively. Although small, means of increases by alternative methods were significantly $(P < .01)$ greater than 0. A considerable number of decreases by alternative methods were seen. Probabilities of decreases in correlations between true and estimated sire breeding values were 8.3, 23.3, 10.0 and 18.3% (5, 14, 6 and 11 incidences from 60 replications) for methods 1, 2, 3 and 4, respectively. In other words, probabilities of increasing accuracy were more than 75% with method 2, 80% with method 4, and 90% with methods 1 and 3. Smaller increases in correlations for sire breeding values were probably due to, and indicative of the greater influence of management groups on breeding values for cows than for sires. Records from multiple daughters of sires are generally much more likely to be a sample representative of population average environment than are records of an individual cow.

Figure 11 summarizes the performance of the four alternative methods by showing mean increases in correlations for management group effects, and for cow and sire breeding values. Mean increases were largest for management group effects for all four alternative methods, and diminished for cow and sire breeding values. Mean increases in correlations for group effects of methods 1 and 2 were smaller than those of methods 3 and 4 probably due to ignoring, in methods 1 and 2, permanent environmental effects of records lacking sire identification. Method 3 produced largest mean increases in correlations for group effects and for breeding values of cows and sires. Method 4 showed some increases but was no better than method 1 in accuracy of breeding values. Method 2 had uniformly least increases in correlations between true and estimated values of all effects.
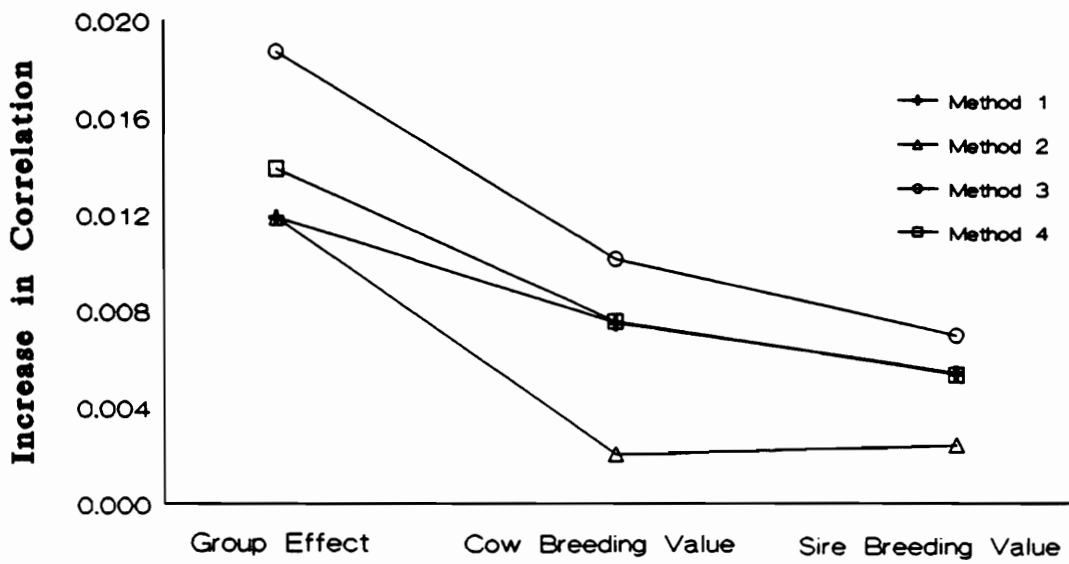
Figure 11. Increases in correlations between true and estimated group effects and breeding values according to each method.

Increases in accuracy of genetic evaluation from the alternative methods examined are expected to occur through increased accuracy of estimating management group effects. Relative increases in accuracy of estimating breeding values can be viewed as a combination of the increase in accuracy of estimating management group effects for each method and the degree to which accuracy of breeding value estimation is associated with more accurate management group effect estimation. Table 7 contains correlations between increases in correlations for management group effects and increases in correlations for cow breeding values (G-C), between increases for group effects and increases for sire breeding values (G-S), and between increases for cow and sire breeding values (C-S).

Accuracy of management group effect estimation was increased most by methods 3 and 4 (Figure 11) which used producing ability to account for differences in genetic merit of cows with unknown sires. However, increases in group effect accuracy were more associated with increased accuracy of breeding value estimation (G-C and G-S, Table 7) for method 3 which treated producing abilities as random effects. The result was an overall superiority of method 3 for accuracy of estimation of group effects and breeding values. In contrast, increased accuracy of breeding values was least associated with increased accuracy of management group estimation for method 2 such that this method produced smallest increases in accuracy for breeding values of cows and sires. This apparently was due to the assumption of method 2 that breeding values of cows lacking sire identity were management group average (as opposed to population average for method 1). Correlations between increases in accuracy of estimated breeding values for sires and cows were approximately 0.6 for all methods except method 2 and reflect the inter-connectedness of evaluations in the animal model.

Table 7. Correlations among increases of correlations between true and estimated group effects, cow and sire breeding values.

Source	Method 1	Method 2	Method 3	Method 4
G-C	.422	.176	.359	.176
G-S	.246	.075	.319	.237
C-S	.605	.510	.608	.628

G-C represents correlations between increases of correlations for group effects and cow breeding values.

G-S represents correlations between increases of correlations for group effects and sire breeding values.

C-S represents correlations between increases of correlations for cow and sire breeding values.

3. Impact of Missing Identifications for Lower Breeding Value Sires

No published evidence of intentional omission of sire identification exists, likely due to the difficulty of collecting reliable information. However, there are circumstances in which intentional omission of sire identification from daughters of sires with known or suspected lower breeding value may be advantageous to individuals proving sires. Additionally, omission of sire identity may be more likely for daughters of natural service sires which have been shown, on average, to be of lower genetic merit than AI sires (77). In these circumstances, use of records lacking sire identification in genetic evaluation may not be advantageous. The objective of this phase of the study was to evaluate the performance of conventional and alternative methods for using records lacking sire identity where sire identity is intentionally omitted for varying proportions of sires with below average breeding values. As shown in c) of Table 1, five populations (replications) of 60 herds with an average per population of 24,917 records and 11,644 cows, were simulated for this phase of the study. Sire identification was randomly deleted as previously to yield, in the first phase, a "high" level of sire identification in which 65% of the cows (7,566 of 11,644) were sire identified. Three sub-populations were formed from the initial population by deleting identification for 50 sires chosen randomly, 25 chosen randomly and 25 lowest true breeding value sires, and 50 lowest true breeding value sires. This produced populations in which identity was intentionally deleted for the lowest 0, 10, and 20% of sires based on true breeding values. The simulated effect was likely larger than in practice since deletions of identity were based on true breeding value and all daughters of each sire were not identified. Average numbers of sire-identified cows after the second phase of deletion were 6186, 6085 and 6045 for 0, 10 and 20% intentional deletion of lowest breeding values, respectively (Table 1). With 20% intentional deletion, breeding values of sires with deleted identity averaged

-715kg. With 10% and 0% intentional deletion, breeding values averaged -410kg and -11kg, respectively.

Table 8 contains means (five populations) of correlations between true and estimated management group effects and breeding values for cows and sires, as well as F-statistics for effects of percentage intentionally deleted sire identity on accuracy of estimated group effects and breeding values. Percentage of intentionally deleted sire identity had no significant effect on accuracy of management group estimates $(P > .05)$ but did affect accuracy of estimating breeding values for cows and sires $(P > .01)$. There was a general trend of reduction in correlations as intentional deletions were increased for all methods, but differences were extremely small for group effects. For breeding values of cows and sires, however, reductions in correlations with increasing percentage intentional deletion were larger $(P < .01)$. As expected, intentional omission of identifications of sires having low breeding values reduced accuracy of genetic evaluation by any method. In the conventional method this reduction in accuracy can be viewed as a selection effect where records subject to selection are not included in the analysis (12, 20).

As shown in Table 9, increases (alternative minus conventional methods) in correlations between true and estimated group effects and breeding values were small but consistently positive except for sire breeding values of method 2 with 20% deletion of sire identity. Increases in correlations for group effects were non-significant $(P > .05)$ with smaller increases at 10% than at 0 or 20% intentional deletion of lower breeding value sires. Increases in accuracy of estimating breeding values of cows increased with larger percent of sire identity deletion, with significant $(P < .05)$ differences due to the proportion of sires deleted in alternative methods 1, 2 and 3. These results suggest alternative methods may be superior to the conventional method for estimating breeding values of cows when identifications of lower breeding value sires are intentionally

Table 8. F-statistics due to proportions of intentionally omitted lower breeding value sires and means of correlations between true and estimated group effects and breeding values.

	Intentional deletion	Conventional method	Method 1	Method 2	Method 3	Method 4
Group effect[a]		.01	.01	.01	.01	.01
	0%	.7302	.7438	.7438	.7502	.7446
	10%	.7282	.7414	.7416	.7480	.7426
	20%	.7270	.7408	.7408	.7470	.7416
Cow breeding value		9.52	7.89	8.89	8.12	8.00
	0%	.6416	.6490	.6430	.6512	.6496
	10%	.6052	.6146	.6090	.6170	.6144
	20%	.5890	.6008	.5960	.6032	.5998
Sire breeding value		9.17	8.42	8.87	8.21	9.34
	0%	.8390	.8426	.8396	.8440	.8440
	10%	.7792	.7884	.7830	.7992	.7890
	20%	.7596	.7658	.7588	.7698	.7664

a Non-significant; others (P < .01).

Table 9. F-statistics due to proportions of intentionally omitted lower breeding value sires and means for increases in correlations between true and estimated group effects and breeding values.

	Intentional deletion	Method 1	method 2	Method 3	Method 4
Group effect		.03	.01	.00	.00
	0%	.0136	.0136	.0200	.0144
	10%	.0132	.0134	.0198	.0144
	20%	.0138	.0138	.0200	.0146
Cow breeding value		5.09[b]	5.26[b]	7.78[b]	3.15
	0%	.0074	.0014	.0096	.0080
	10%	.0094	.0038	.0118	.0092
	20%	.0118	.0070	.0142	.0108
Sire breeding value		1.99	.98	2.06	1.16
	0%	.0036	.0006	.0054	.0054
	10%	.0092	.0038	.0130	.0098
	20%	.0062	-.0008	.0102	.0068

b Significant ($P < .05$); others ($P > .05$).

omitted. Increases in correlations between true and estimated breeding values of sires were not affected significantly ($P > .05$) by differences in proportion of lower breeding value sires deleted intentionally. Increases in correlations for sires generally were positive, less than those for cows, and without noticeable tendency to increase with increasing proportion of sire identity deleted. As previously, largest increases in correlations tended to be for method 3. Unlike previous results (Figures 7, 11, 12 and 14), increases in correlations for sire breeding values by method 4 were higher than those by method 1 with intentional omission of identity for lower breeding value sires. The assumption of method 1 that genetic values of records lacking sire identification were population genetic average logically is less appropriate where the absence of sire identity is related to sire breeding value. Methods 3 and 4, which compute genetic value of cows lacking sire identification from cow performance, are expected to be superior under these circumstances.

4. Performance of Alternative Methods in Individual Herds

The relative accuracy of alternative methods for incorporating records lacking sire identity into genetic evaluation programs will be affected by herd size, proportion of sire identified cows, and the level of variation in temporary herd effects. Herd size and proportion of sire identified cows determine the size (number of records) of management groups and/or the homogeneity of their effects and therefore the accuracy of estimation and/or efficiency of removing environmental group effects from the genetic evaluation. Variation in temporary herd effects partially determines the level of variation among management groups and consequently the importance of accurately accounting for differences among management groups. To examine the effects of these factors on the accuracy of alternative methods, genetic evaluations by conventional

and alternative methods were computed from simulation involving three herd sizes (40, 70 and 100), three proportions of sire identified cows (20, 50 and 80%), and two standard deviations for temporary herd effects (552kg and 733kg). Different numbers of herds were used for each herd size (80 herds for herd size 40, 60 herds for herd size 70, and 40 herds for herd size 100), and averages of 7,279 cows, 9,628 cows and 9,175 cows resulted. Deletion of sire identification by comparing random numbers (ranged from 0 to 1) assigned to cow with given probabilities (.2, .5 and .8) yielded proportions of sire identified cows nearly equal to those planned (Table 1).

Herd Size and Proportion of Sire-Identified Cows

Table 10 contains ratios of numbers of groups (R1-R4) and average numbers of records in groups (M1-M4) in simulated populations. A total of 80, 60 and 40 herds were simulated for herd size 40, 70 and 100, respectively, to maintain similar numbers of cows. As a result, the number of herd-years differed for each herd size. In addition, numbers of sires simulated varied for different combinations of factors to maintain nearly equal numbers of identified daughters per sire (Table 1). As shown in previous simulations, numbers of herd-year-seasons were equal for all alternative methods, while numbers of management groups were equal for methods 1 and 2 and for methods 3 and 4. Average numbers of management groups in methods 3 and 4, which were slightly larger than methods 1 and 2, were used to compute ratios R2 and R4 in Table 10.

The value of R1 indicates the relative number of herd-year-seasons with one or more records usable by alternative and conventional methods, while R2 is the ratio of numbers of management groups eventually formed (5 or more records) in alternative and conventional methods. Both alternative and conventional genetic evaluation procedures were subjected to equal levels of environmental variation from the same num-

Table 10. Ratio among numbers of herd-year-season and management groups, and average size of groups.

Herd size	Sire identified proportion	R1	R2	R3	R4	M1	M2	M3	M4
	20%	1.40	4.06	.28	.80	1.76	6.43	6.43	8.10
40	50%	1.04	1.75	.48	.81	3.41	7.08	6.63	8.20
	80%	1.01	1.17	.68	.80	5.21	7.62	6.50	8.17
	20%	1.12	2.97	.37	.99	2.46	6.85	11.50	11.67
70	50%	1.01	1.35	.73	.98	5.70	7.80	11.46	11.65
	80%	1.00	1.05	.94	.99	9.12	9.68	11.41	11.59
	20%	1.04	2.14	.49	1.00	3.43	7.05	16.45	16.46
100	50%	1.00	1.10	.91	1.00	8.17	8.99	16.34	16.35
	80%	1.00	1.01	.99	1.00	13.07	13.15	16.38	16.41

R1 = ratio of number of herd-year-seasons in alternative methods to
number of herd-year-seasons in conventional method.
R2 = ratio of number of management groups in alternative methods 3 and 4
to number of management groups in conventional method.
R3 = ratio of number of management groups to
number of herd-year-seasons in conventional method.
R4 = ratio of number of management groups to number of
herd-year-seasons in alternative methods 3 and 4.
M1 = average size of herd-year-seasonal group in conventional method.
M2 = average size of management group in conventional method.
M3 = average size of herd-year-seasonal group in alternative methods.
M4 = average size of management group in alternative methods 3 and 4.

ber of herd-years, but differences existed in the numbers of potential management groups (herd-year-seasons) for alternative and conventional methods as shown by R1. Results indicated that from zero to 40 percent more herd-year-seasons contained at least one record usable by alternative methods, while the number of management groups formed with alternative methods (R2) ranged from approximately equal to slightly more than four times the number formed using the conventional method of discarding records lacking sire identification. As expected, both ratios were higher with smaller herd size and lower percentages of sire identification.

The ratios R3 and R4 in Table 10 reflect the percentage of herd-year-seasons which were directly usable as management groups for conventional and alternative methods. Conversely, R3 and R4 are related inversely to the frequency with which it was necessary to combine adjacent herd-year-seasons to obtain management groups of five or more records. Unit values of R3 and R4 indicate no combining of herd year seasons to form management groups. With herd size of 100 and 80% sire identification, nearly all herd-year-seasons contained sufficient numbers of records to form a management group with at least five records using the conventional method (R3 = .99). Results were similar for herd size 100 and 50% sire identified (R3 = .91) and herd size 70 with 80% sire identified (R3 = .94). For other combinations of herd size and percentage sire identification, there was substantially more need to combine herd-year-seasons to obtain management groups of acceptable size with the conventional method (R3 = .73 to .28). In contrast, the proportion of herd-year-seasons which were used to form management groups with alternative methods 3 and 4 was never less than .80 (R4).

Values for M1 and M2 in Table 10 indicate average numbers of records in herd-year-seasons and management groups for the conventional method. Equivalent results are shown for alternative methods by M3 and M4. Average number of records in

herd-year-seasons usable in the conventional method (i.e., sire identified) ranged from 1.76 with herd size 40 and 20% sire identified to 13.07 with herd size 100 and 80% sire identified. For four of the nine situations simulated, the average number of records per herd-year-season (M1) was less than the five acceptable for management group formation. In contrast, average records per herd-year-season exceeded five for all simulations using alternative methods (M3). Average size of both herd-year-season and management groups were larger for alternative methods with their difference in average size increasing as herd size increased. A reduced need in alternative methods to combine herd-year-season groups to form management groups should more nearly conform to the model assumption of a common environmental effect for all animals in each group, while larger average group size should yield more accurate estimation of the group effect.

Table 11 contains F-statistics for effects of variation in herd size and proportion of sire identified cows on increases in correlations (alternative minus conventional methods) between true and estimated management group effects and breeding values of cows and sires. Increases in correlations are shown by herd size and proportion of records sire identified in Figure 12 and 13, respectively. Differences in herd size significantly ($P < .05$) affected increases in correlations between true and estimated group effects and cow breeding values (except with method 4), but not for sire breeding values. All alternative methods increased slightly the accuracy of estimation for all effects.

Increases in correlations for group effects by alternative methods were greater at larger herd sizes (Figure 12). This was likely due to the larger increase in herd-year-season group size with alternative methods in larger herds (M1 vs. M3 in Table 10). For example, with herd size 40, there were between 1.2 and 3.7 times as many cows in herd-year-season groups with alternative methods as with the conventional method. At herd size 100, these values ranged from 1.3 to 4.8. Also, average management group

Table 11. F-statistics due to herd size, proportion of sire-identified cows and variation in temporary herd effects for increases in correlations between true and estimated breeding values.

Estimated effect	Source	df	Method 1	Method 2	Method 3	Method 4
Management group	Herd size	2	10.53	10.53	4.14	6.93
	Proportion	2	47.31	47.31	11.34	.64[a]
	Variation[a]	1	.71	.71	.32	.16
	HxP[a]	4	6.47	6.47	3.77	3.83
	HxV[a]	2	.44	.44	1.67	.58
	PxV[a]	2	.91	.91	.34	.48
	HxPxV[a]	4	1.16	1.16	1.06	.24
Cow breeding value	Herd size	2	12.51	4.85	16.26	2.60[a]
	Proportion	2	212.55	65.82	190.24	51.91
	Variation	1	20.92	14.71	23.34	13.91
	HxP[a]	4	1.05	1.12	.56	.56
	HxV[a]	2	.51	.67	.42	.27
	PxV	2	4.63	2.62[a]	4.89	3.51
	HxPxV[a]	4	1.18	1.23	1.20	1.15
Sire breeding value	Herd size[a]	2	1.37	1.21	1.09	.33
	Proportion	2	31.50	24.65	18.13	6.79
	Variation	1	4.12	2.55[a]	2.99[a]	3.20[a]
	HxP	4	1.61[a]	2.89	.95[a]	.31[a]
	HxV[a]	2	.94	.98	.66	.56
	PxV[a]	2	1.27	.39	.95	1.46
	HxPxV[a]	4	.24	.28	.26	.18
	Error	36				
	Total	53				

a Non-significant; others ($P < .05$)

Figure 12. Increases in correlations between true and estimated group effects and breeding values by alternative methods according to different herd sizes.

size (M2 and M4) with alternative methods ranged from .55 to 1.67 cows larger than the conventional method with herd size 40. Comparable values for herd size 100 were 3.34 to 9.41 cows. Less combining of management groups and larger numbers of records in management groups should increase the accuracy with which the group effect is estimated. Correlations with alternative methods (.6912, .7167 and .7172 for methods 1 and 2 across herd size, .6978, .7207 and .7231 for method 3, and .6806, .7151 and .7173 for method 4) increased as herd size increased. However, correlations with the conventional method increased less (.6778, .7006 and .6931 for herd sizes 40, 70 and 100). Consequently increases for management group effects were greatest at herd size 100.

Increases in correlations between true and estimated breeding values of cows decreased as herd size increased (Figure 12). This was expected due to the relatively higher accuracy of the conventional method (.6060, .6210 and .6293 for herd size 40, 70 and 100) as herd size increased. Accuracy of estimated cow breeding values by the conventional method were greatest at herd size 100 with correspondingly less opportunity for improvement by alternative methods. That is, correlations by alternative methods increased slightly with herd size (e.g. for method 3; .6367, .6460, and .6482 for herd sizes 40, 70 and 100, respectively), but less than correlations by the conventional method.

Accuracies of estimating breeding values of sires were increased by all alternative methods but were not significantly ($P > .05$) affected by differences in herd size. Accuracy of sire evaluation is determined mostly by numbers of daughters which was held nearly constant in simulations of different herd sizes. Increases in accuracy for estimated breeding values of sires ranged from about .015 to .020 in contrast with ranges of approximately .020 to .030 for group effects and cow breeding values. Alternative

method 3 consistently yielded the greatest increases in accuracy for estimating all effects.

In the conventional method of estimating breeding value, a low proportion of sire-identified cows causes the loss of a high proportion of data and frequent combining of herd-year-seasons to form management groups. In contrast, records lacking sire identification were considered as records of group mates in all alternative methods, such that the frequency of combining groups did not depend on the percentage of sire-identified cows. If there was no record with sire identification in a combined management group, group effects were not estimated in methods 1 and 2, but methods 3 and 4 used records lacking sire identity to estimate producing abilities. F-statistics in Table 11 for effects of percentage of sire identified cows on increases of correlations indicated that improvements in accuracy of estimation for all effects were more dependent on the percentage of sire-identified cows than on herd size (except group effects by method 4). When the portion of records lacking sire identification was large, greatest improvements in accuracy were obtained by alternative methods (Figure 13). The conventional method required frequent combining of herd-year-seasons to form management groups and yielded substantially smaller average group size (M2, Table 10) when the percentage of sire-identified cows was low. Advantages of alternative methods were greatest at low percent of sire-identified cows, since a high proportion of records conventionally discarded were used to remove management group effects from genetic evaluations. As seen previously, estimation accuracy of alternative method 3 generally was highest for all effects although method 1 was essentially equivalent at 20 percent sire identified records.
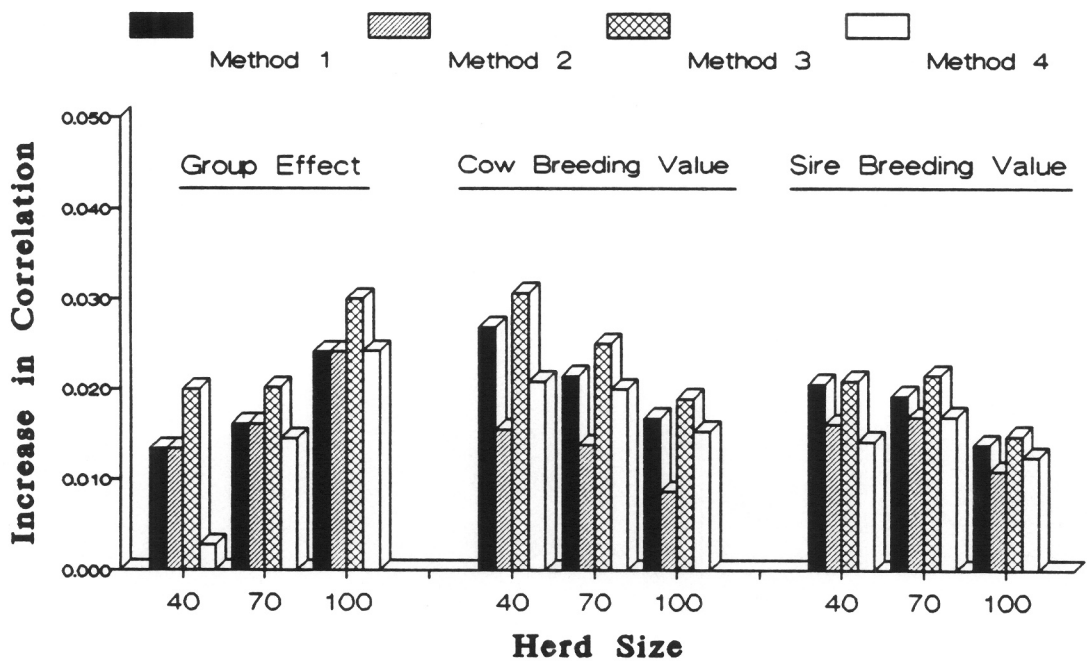
Figure 13. Increases in correlations between true and estimated group effects and breeding values by alternative methods according to different proportions of sire-identified cows

Variation in Temporary Herd Effects

The importance of effectively removing environmental differences from genetic evaluations depends on the amount of variation in performance records which results from differences in environmental conditions. Temporary herd effects contribute to management group effects such that an increase in variance of temporary herd effects will increase management group variance. In previous simulations, temporary herd effects were generated randomly from a normal distribution ($\mu = 0$; $\sigma_{th} = 552$kg) and assigned to all cows freshening in a common herd, year and 2 month season. To investigate the effect of increasing environmental variation on accuracy of conventional and alternative methods, the variation in temporary herd effect was increased (552 vs 733kg in standard deviation).

Figure 14 shows increases in correlations between true and estimated group effects and breeding values by alternative methods under two different levels of variation in temporary herd effects. Increases in correlations for group effects were slightly smaller for the high standard deviation (733kg) probably due to larger variation in management groups of both conventional and alternative methods than for the low (552kg). However, differences due to level of variation were very small and not significant (P > .05; Table 11). Differences in increases in correlations between true and estimated breeding values due to variation in temporary herd effects were significant (P < .05) for cows except for method 2 (Table 11). However, differences were non-significant (P > .05) for sire breeding values except for method 1. As expected, the accuracy of alternative methods increased slightly relative to the conventional method as variation due to temporary herd effects was increased. Alternative methods are not expected to estimate group effects more accurately with more variation due to group effects. However, any superiority in estimation accuracy (i.e., with any level of variation) is
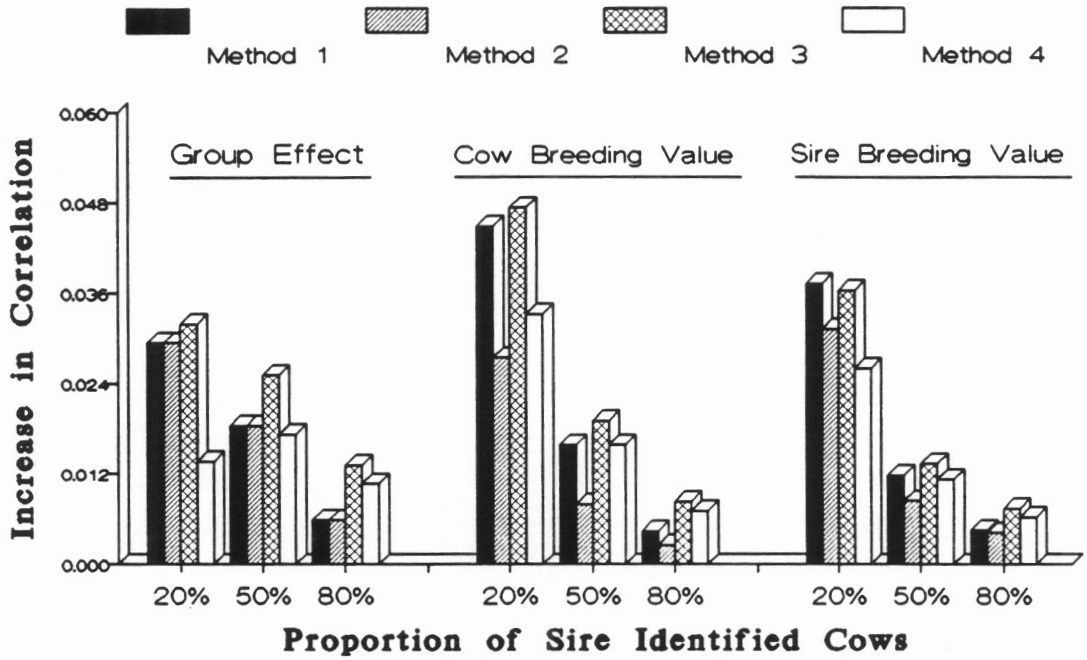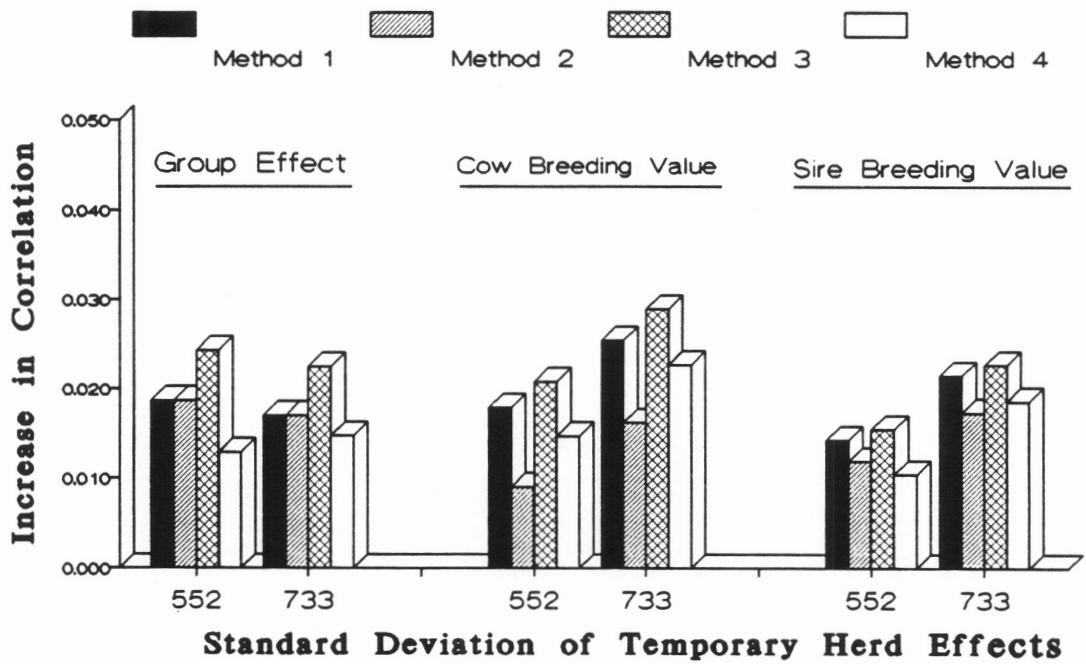
Figure 14. Increases in correlations between true and estimated group effects and breeding values by alternative methods under two different variation of temporary herd effects.

expected to have larger impact on genetic evaluation as non-genetic variation increases. Relative to the conventional method, accuracies were increased between .015 and .030 with method 3 generally superior to other methods.

The interaction between proportion of sire identified cows and variation in temporary herd effects was significant ($P < .05$) for cow breeding values except for method 2. In general, however, interactions did not significantly ($P > .05$) affect increases in correlations for any estimated effects.

Computing Time

An important factor in comparing alternative methods of genetic evaluation is computing time, which was mostly dependent on percentage of sire-identified cows with the conjugate gradient method employed in this study. As shown in Figure 15, computing times of methods 1 and 2 were approximately equal to that of the conventional method. However, methods 3 and 4 required more than 3× and nearly 5×, respectively, the computing time of the conventional method with 20% of sire-identified cows. Computing time for alternative methods at 80% of sire-identified cows was nearly the same as that for the conventional method. Simultaneous consideration of improvements in accuracy and computing time might suggest the use of method 3 for herds with high, and method 1 for herds with low, percentage of sire-identified cows. However, computing time would be different depending on the algorithms used to solve the mixed linear model equations. As mentioned before, the conjugated gradient method for a sparse linear system was employed in this study. Thus, computing time was determined by number of non-zero elements (Table 2) and the condition-number derived from equation [2.24]. Comparison of computing time in Figure 15 may not be applicable to genetic evaluations using other algorithms.
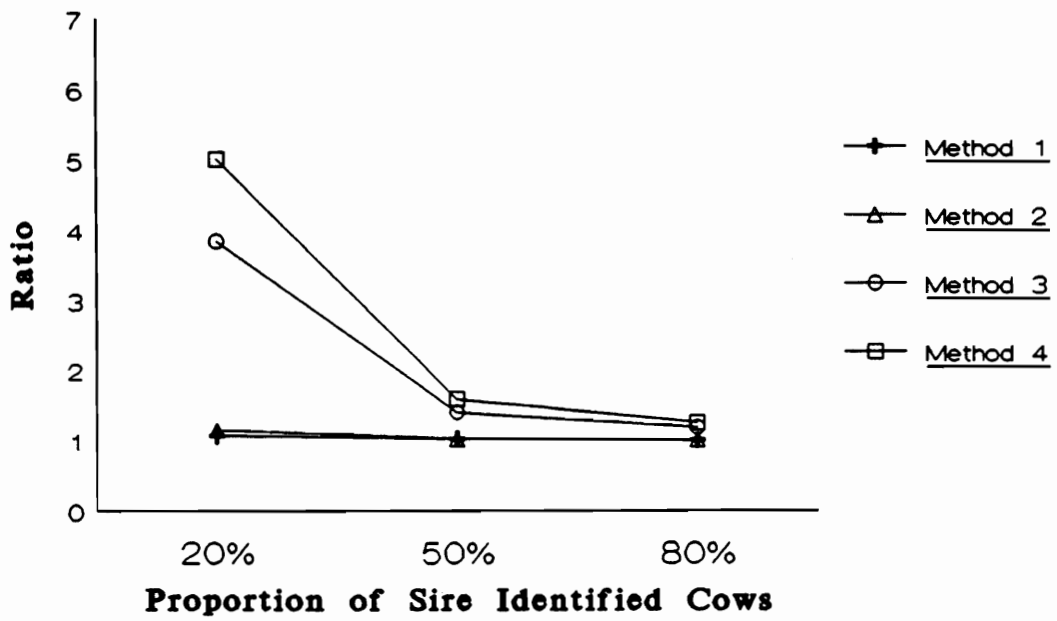
Figure 15. Computing time ratio of alternative methods to the conventional method.

CONCLUSIONS

Management groups are defined in genetic evaluation to remove effects of environmental factors common to individuals within groups. Records of cows lacking sire identity contain information potentially useful in estimating effects of management groups. However, differences in average genetic value among management groups must be accounted for to avoid biases in estimation of the group environmental effects. In the USDA animal model, all pedigree paths end with animals having unknown parents whose genetic values are assumed equal to the average genetic value of individuals in the genetic group to which they are assigned. The genetic value of cows lacking sire or parent identity can be considered as analogous.

In this study, four alternative methods were used to account for genetic values of cows lacking sire identity. Method 1 assumed cows lacking sire identity had genetic value equal to the average animal born in the population evaluated. With actual data, the population used for method 1 likely would include geographic region and registration status, which were not simulated in this study. In method 2, genetic values for cows lacking sire identification were assumed to equal the average genetic value for sire identified cows in their management group. Permanent environmental effects were ignored in both methods 1 and 2. Method 2 would be expected to produce substantially more genetic (or apparent genetic) variation among management groups than method 1, and potentially to assign to many cows, the average estimated breeding value of a few (for adjustment purposes only). Although differences were small, method 2 tended to yield estimates of cow and sire breeding value that were less accurate than method 1.

Methods 3 and 4 adjusted estimates of management group effects for genetic and permanent environmental effects by estimating the producing abilities of cows from their own records. Method 3 estimated producing abilities as random effects while producing abilities were considered fixed in method 4. In theory, method 3 should be optimum among the alternatives examined since both genetic and permanent environmental effects are considered and treated as random variables. The major shortcoming of method 3 and 4 in practice would be high prediction error variance for producing abilities due to limited numbers of records on individual cows. In spite of this potential shortcoming, results of this study (with an average of approximately 2.1 records per cow) showed method 3 to estimate effects of management groups and breeding values of cows and sires more accurately than the conventional or any other alternative method. Although differences were small, method 3 was superior under all situations examined.

The effect of minimum management group size on accuracy of estimation differed for estimation of group effects and breeding values. As expected, group effects were estimated more accurately with large group size. However, breeding values were more accurately estimated with smaller sized groups which were both more numerous and more homogeneous. Since the goal of genetic evaluation is accurate estimation of breeding values, the USDA animal model minimum group size of five (or fewer) animals would seem, based on results of this study, to be a near optimum value.

Omission of sire identity from daughters of lower breeding value sire reduced accuracy of breeding value estimation for conventional and all alternative methods. Where identity was omitted on the 10 and 20 percent of sires with lowest breeding value, correlations between true and estimated breeding values were reduced approximately .035 and .055 for cows and sires, respectively at 10%, and .049 and .078, respectively at 20%. Average accuracy of estimating breeding value generally was higher

for alternative than conventional methods (one exception was for sire breeding values by method 2 at 20% omission). Results indicate that breeding values estimated by alternative methods are not more susceptible to biases from intentional omissions of breeding values for sires of lower breeding value.

It may be possible to apply different alternative methods to individual herds by adopting different subroutines for incorporating records lacking sire identity according to herd characteristics. Percentage of sire identified cows appeared to be the most important characteristic determining the improvement in accuracy by alternative methods. Correlations between all true and estimated effects were uniformly higher for alternative methods relative to the conventional method when the proportion of sire identified cows was low. With 20% sire identified records, correlations were increased approximately .045 and .038 for cow and sire breeding values, respectively, by alternative methods. Herd size also had a significant (P < .05) effect on increases for alternative methods in accuracy of estimating breeding values of cows (except for method 4), but not for sires. Differences in accuracy between conventional and alternative methods due to the levels of variation examined in this study were seldom significant (P > .05). In general, results indicated that alternative methods may be used most effectively in smaller herds with low proportions of sire identified records and higher within-herd variance.

All simulations studies are limited to some extent by an inability to duplicate practical conditions exactly and by the range of parameters examined. Several characteristics of the simulation used in present study should be considered for possible effect on results obtained and interpretations made.

First, no genetic differences were simulated across herds or management groups. Alternative method 2, which assumed genetic values of cows with missing sire identification to equal average genetic value of sire identified cows in the same management

group, generally performed less well than methods 1 or 3. Large genetic differences among management groups logically would have made method 2 relatively more attractive.

Additionally, no auto-correlated error structure was simulated among management groups. Although mutually uncorrelated errors are assumed by all genetic evaluation methods, it seems more reasonable to suppose that, in reality, management groups adjacent in time or location are more alike than those widely separated in either or both dimension. The lack of correlation between adjacent management groups in simulation may have exaggerated consequences of combining management groups to achieve minimum group size.

Lastly, it should be denoted that both sires and the occurrence of missing sire identification were distributed randomly across management groups in these simulated data. The former should favor accurate genetic evaluation for all methods with perhaps relatively little effect on comparisons among conventional and various alternative methods. The effect of uniform distribution across herds of missing sire identification, whether random or for low breeding value sires, is unknown but may not represent a worst case scenario with respect to consequences for accuracy of genetic evaluation.

Computing time is an important consideration in genetic evaluation of large population. Alternative method 3 without any assumption for genetic values of records lacking sire information consistently yielded estimates of management group effects and breeding values which were more accurate than the conventional method, and most accurate of the alternatives examined. However, method 3 required nearly four times the computing time of the conventional method and methods 1 and 2 with the lowest proportion of sire identification, where alternative methods should be most useful. Since accuracies of alternative methods 1 and 3 are similar, method 1 may be

the optimum method for incorporating records lacking sire identity, given computing algorithm used in this study.

REFERENCES

1. Akima, H. 1970. A new method of interpolation and smooth curve fitting based on local procedures. J. Assoc. Comp. Machinery, 17:589.

2. Allaire, F. R., and S. N. Gaunt. 1965. First lactation contemporary comparisons as indicators of environmental influences on daughter records used for sire evaluation. J. Dairy Sci. 48:454.

3. Bereskin B., and A. E. Freeman. 1965. Genetic and environmental factors in dairy sire evaluation. I. Effects of herds months, and year-seasons on variance among lactation records; repeatability and heritability. J. Dairy Sci. 48:347.

4. Butcher, K. R., and J. E. Legates. 1976. Estimating son's progeny test from his pedigree information. J. Dairy Sci. 59:137.

5. Cassell, B. G., and B. T. McDaniel. 1983. Use of later records in dairy sire evaluation: A review. J. Dairy Sci. 66:1.

6. Dickinson, F. N., H. D. Norman, J. F. Keown, and L. G. Waite. 1974. Revisions to USDA methodology for sire summaries and cow indexes. J. Dairy Sci. 57:977.

7. Dickinson, F. N., R. L. Powell, H. D. Norman, L. G. Waite, and B. T. McDaniel. 1976. The USDA-DHIA modified contemporary comparison sire summary and cow index procedures. USDA-ARS Prod. Res. Rep. 165.

8. Fairchild, T. P., W. J. Tylor, G. R. Barr, and E. L. Coley. 1966. Estimating transmitting abilities of artificial insemination dairy sires. J. Dairy Sci. 49:1416.

9. Falconer, D. C. 1981. Introduction to quantitative genetics. John Wiley & Son Inc. New York.

10. Foulley, J. L., D. Gianola, and D. Planchenault. 1987. Sire evaluation with uncertain paternity. Genet. Sel. Evol., 19:83.

11. Freeman, A. E. 1973. Age adjustment of production records; history and basic problems. J. Dairy Sci. 56:941.

12. Gianola, D., and R. L. Fernando. 1986. Bayesian methods in animal breeding theory. J. Animal Sci. 63:217.

13. Hargrove, G. L., and J. E. Legates. 1971. Biases in dairy sire evaluation attributable to genetic trend and female selection J. Dairy Sci. 54:1041.

14. Hargrove, G. L., H. W. Thoele, R. N. Deb and J. L. Gobble. 1974. Sire evaluation methods. J. Dairy Sci. 57:889.

15. Heidues, T., L. D. Van Vleck, and C. R. Henderson. 1961 Actual and expected accuracy of sire proofs under the New York system of sampling bulls. Z. Tierzuchsbiol. 75:323.

16. Henderson, C. R. 1949. Estimation of changes in herd environment. J. Dairy Sci. 32:706 (Abstr.).

17. Henderson, C. R., H. W. Carter, and J. T. Godfrey. 1954. Use of contemporary herd average in appraising progeny tests of dairy bulls. J. Animal Sci. 13:959.

18. Henderson, C. R., O. Kempthorne, S. R. Searle, and C. M. von Krosigk. 1959. The estimation of environmental and genetic trends from records subject to culling. Biometrics 15:192-218.

19. Henderson, C. R. 1963. Selection index and expected genetic advance. Page 141 in: Statistical genetics and plant breeding. Hanson, W. D., and H. F. Robinson (eds). NAS-NRC, Pub. 982.

20. Henderson, C. R. 1973. Sire evaluation and genetic trends. In Proceeding of the Animal Breeding and Genetics Symposium in Honor of Dr. J.L. Lush. ASAS and ADSA, Champaign Il.

21. Henderson, C. R. 1974. General flexibility of linear model techniques for sire evaluation. J. Dairy Sci. 57:963.

22. Henderson, C. R. 1975. Use of relationships among sires to increase accuracy of sire evaluation. J. Dairy Sci. 58:1731.

23. Henderson, C. R. 1975. Use of all relatives in intraherd prediction of breeding values and producing abilities. J. Dairy Sci. 58:1910.

24. Henderson, C. R. 1975. Comparison of alternative sire evaluation methods. J. Animal Sci. 41:760.

25. 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.

26. Henderson, C. R. 1978. Simulation to examine distribution of estimators of variances and ratios of variances. J. Dairy Sci. 61:267.

27. Hickman, C. G., and C. R. Henderson. 1955. Component of the relationship between level of production and rate of maturity in dairy cattle. J. Dairy Sci. 38:883.

28. Hickman, C. G. 1973. Herd-level methods for age adjustment of milk yields. J. Dairy Sci. 56:947.

29. IMSL user's manual Stat/Library. 1987. Houston, Texas.

30. Johansson, I., and A. Hansson. 1940. Causes of variation in milk and butter fat yield of dairy cows. K. Lantbr. Akad. Handb. 79:1.

31. Johansson, I. 1961. Genetic aspects of dairy cattle breeding. University of Illinois Press. Urbana.

32. Keown, J. F. 1974. Comparison of mixed model methods of sire evaluation. J. Dairy Sci. 57:245.

33. Kennedy, B. W. 1981. Bias and mean square error from ignoring genetic groups in mixed model sire evaluation. J. Dairy Sci. 64:689.

34. Kennedy, B. W., L. R. Shaeffer, and D. A. Sorensen. 1988. Genetic properties of animal models. J. Dairy Sci. 71(Suppl. 2):17.

35. Lawlor, R. L., P. M. VanRaden, T. H. Short and I. Hoeschele. 1991. Evaluation of identical animals and full sibs by an animal model including dominance relationships. J. Dairy Sci. 74(Suppl. 1):156 (Abstr.).

36. Lee, A. J. 1974. Month, year, and herd effects on age adjustment of first lactation milk yield. J. Dairy Sci. 57:332.

37. Lee, A. J. 1976. Relationship between milk yield and age at calving in first lactation. J. Dairy Sci. 59:1794.

38. Lentz, W. E., P. D. Miller, and C. R. Henderson. 1969. A new method of dairy sire evaluation. J. Animal Sci. 29:109 (Abstr.).

39. Lin, C. Y. 1988. Four equivalent sets of mixed-model equations with relationship matrix for estimation of genetic parameters. J. Animal Sci. 66:1627.

40. Lush, J. L. 1947. Family merit and individual merit as bases for selection. Part I. The American Naturalist. 82:367.

41. Lush, J. L., and R. R. Shrode. 1950. Changes in milk production with age and milking frequency. J. Dairy Sci. 33:338.

42. Mao, I. L., J. W. Wilton, and E. B. Burnside. 1974. Parity in age adjustment for milk yield and fat yield. J. Dairy Sci. 57:100.

43. McDaniel, B. T. 1973. Merits and problems of adjusting to other than mature age. J. Dairy Sci. 56:959.

44. McDaniel, B. T., H. D. Norman, and F. N. Dickinson. 1973. Herdmates versus contemporaries for evaluating progeny tests of dairy bulls. J. Dairy Sci. 56:1545.

45. McDaniel, B. T. 1974. Why new sire summaries are needed. J. Dairy Sci. 57:951.

46. McDaniel, B. T., H. D. Norman, and F. N. Dickinson. 1974. Variation in genetic merit of sires of herdmates of first lactation cows. J. Dairy Sci. 57:1234.

47. McGilliard, M., and D. Edlund. 1978. Dairy cattle breeding simulation. Unpublished mimeograph.

48. Mendenhall, W., R. L. Scheaffer, and D. D. Wackerly. 1986. Mathematical statistics with applications. Third edition. PWS Publisher, Massachusetts.

49. Meyer, K., and W. G. Hill. 1983. A note on the effects of sampling errors on the accuracy of genetic selection indices. Z. Tierzuchtg. Zuchtgsbiol 100,27.

50. Meyer, K. 1987. Variance due to sire x herd interactions and common environmental covariances between paternal half sisters for first lactation production. J. Dairy Sci. 70(Suppl. 1):126 (Abstr.).

51. Meyer, K. 1989. Approximate accuracy of genetic evaluation under an animal model. Livest. Prod. Sci. 21:87.

52. Miller, P. 1973. A recent study of age adjustment. J. Dairy Sci. 56:952.

53. Miller, P. D., W. E. Lentz and C. R. Henderson. 1970. Joint influence of month and age of calving on milk yield of Holstein cows in the northeastern United States. J. Dairy Sci. 35:351.

54. Miller R. H., B. T. McDaniel, and R. D. Plowman. 1968. Comparison of three methods of sire evaluation. J. Dairy Sci. 51:782.

55. Misztal, I., and G. R. Wiggans. 1988. Approximation of prediction error variances in large-scale animal models. J. Dairy Sci. 71(Suppl.2):27.

56. Norman, H. D., B. T. McDaniel, and F. N. Dickinson. 1972. Conflicts between heritability estimates of mature equivalent and herdmate-deviation milk and fat. J. Dairy Sci. 55:107.

57. Norman, H. D., B. T. McDaniel, and F. N. Dickinson. 1972. Regression of daughter and herdmate milk yield on genetic value of the herdmates' sires. J. Dairy Sci. 55:1735.

58. Norman, H. D. 1974. Factors that should be considered in a national sire summary model. J. Dairy Sci. 57:955.

59. Norman, H. D., and R. L. Powell. 1983. Use of grade animals in United States Department of Agriculture Programs. J. Dairy Sci. 66:1567.

60. Norman, H. D., G. R. Wiggans, C. A. Ernst, and J. R. Wright. 1989. Determination of size of management group for optimizing predictability in the animal model. J. Dairy Sci. 72(Suppl. 1):71 (Abstr.).

61. Norman, H. D., R. L. Powell, and G. R. Wiggans. 1991. Comparison of genetic evaluations from animal model and modified contemporary comparison. J. Dairy Sci, 74:2309.

62. Pearson, R. E. 1971. The effect of age distribution and female culling on the profitability of dairy herd. Ph.D. Thesis. Iowa State University, Ames.

63. Plowman, R. D., and B. T. McDaniel. 1968. Changes in USDA sire summaries procedures. J. Dairy Sci. 51:306.

64. Pollak, E. J., and R. L. Quaas. 1983. Definition of group effects in sire evaluation models. J. Dairy Sci. 66:1503.

65. Powell, R. L., and A. E. Freeman 1974. Estimates of sire merit. J. Dairy Sci. 57:1228.

66. Powell, R. L., H. D. Norman, and F. N. Dickinson. 1977. Relationships between bulls' pedigree indexes and daughter performance in modified contemporary comparison. J. Dairy Sci. 60:961.

67. Powell, R. L. 1978. A procedure for including the dam and maternal grandsire in USDA-DHIA cow indexes. J. Dairy Sci. 61:794.

68. Powell, R. L., and H. D. Norman. 1981. Where are the best cows ? Hoard's Dairyman 126:1315.

69. Powell, R. L. 1985. Cow evaluation procedures. National Cooperative Dairy Herd Improvement Fact Sheet. H-2.

70. Powell, R. L., and H. D. Norman. 1986. Genetic and environmental differences between registered and grade Holstein cows. J. Dairy Sci. 69:2897.

71. Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. 1989. Numerical recipes. Cambridge University Press.

72. Quaas, R. L., R. W. Everett, and A. C. McClintock. 1979. Maternal grandsire models for dairy sire evaluation. J. Dairy Sci. 62:1648.

73. Quaas, R. L., and E. J. Pollak. 1980. Mixed model methodology for farm and ranch beef cattle testing programs. J. Animal Sci. 51:1277.

74. Quaas, R. L., and E. J. Pollak. 1981. Modified equations for sire models with groups. J. Dairy Sci. 64:1868.

75. Quaas, R. L., R. D. Anderson, and A. R. Gilmore. 1984. BLUP school handbook. University of New England.

76. Quaas, R. L. 1988. Additive genetic model with groups and relationships. J. Dairy Sci. 71:1388.

77. Robertson, A., and J. M. Rendel. 1954. The performance of heifers got by artificial insemination. J. Agr. Sci. 44:184.

78. Rohatgi, V. K. 1976. An introduction to probability theory and mathematical statistics. John Wiley & Sons, Inc.

79. Ronningen, K. 1974. Monte Carlo simulation of statistical-biological models which are of interest in animal breeding. Acta Agriculture Scandinavica 24:135.

80. Sanders, H. G. 1928. The variations in milk yields caused by season of the year, service, age, and dry period, and their elimination. J. Agr. Sci. 18:46.

81. Schaeffer, L. R. 1983. Notes on linear model theory, best linear unbiased prediction, and variance components. Guelph university, Guelph, Canada.

82. Searle, S. R., and C. R. Henderson. 1960. Judging the effectiveness of age-correction factors. J. Dairy Sci. 43:966.

83. Searle, S. R. 1962. Age and herd effects in New Zealand dairy cow records. J. Dairy Sci. 45:82.

84. Thompson, R. 1976. Relationship between the cumulative difference and best linear unbiased predictor methods of evaluating bulls. Anim. Prod. 23:15.

85. Thompson, R. 1979. Sire evaluation. Biometrics 35:339.

86. VanRaden, P. M., and G. R. Wiggans. 1991. Derivation, calculation, and use of national animal model evaluation. J. Dairy Sci. 74:2737.

87. Van Vleck, L. D., Heidhues, and C. R. Henderson. 1961. Analysis of derivations of dairy records from different contemporary averages. J. Dairy Sci. 44:269.

88. Van Vleck, L. D. 1963. Regression of records on herd-mate averages. J. Dairy Sci. 46:846.

89. Van Vleck, L. D. 1987. Contemporary groups for genetic evaluations. J. Dairy Sci. 70:2456.

90. Vinson W. E. 1987. Potential bias in genetic evaluations from differences in variation within herds. J. Dairy Sci. 70:2450.

91. Westell, R. A., and L. D. Van Vleck. 1987. Simultaneous genetic evaluation of sires and cows for a large population of dairy cattle. J. Dairy Sci. 70:1006.

92. Westell, R. A., R. L. Quass, and L. D. Van Vleck. 1988. Genetic groups in an animal model. J. Dairy Sci. 71:1310.

93. White, J. M., W. E. Vinson, and R. E. Pearson. 1981. Dairy cattle improvement and genetics. J. Dairy Sci. 64:1305.

94. Wickham, B. W., and C. R. Henderson. 1977. Sire evaluation by second lactation records of daughters. J. Dairy Sci. 67:893.

95. Wiggans, G. R., and D. R. Ostrander. 1984. Counts of 305-day lactation records excluded from genetic evaluation for 1983. National Cooperative Dairy Herd Improvement Program. Fact Sheet. H-2.

96. Wiggans, G. R., and R. L. Powell. 1984. Increasing pedigree contribution to dairy sire evaluation. J. Dairy Sci. 67:893.

97. Wiggans, G. R., I. Misztal, and L. D. Van Vleck. 1988. Implementation of an animal model for genetic evaluation of dairy cattle in the United States. J. Dairy Sci. 71(Suppl. 2):54.

98. Wiggans, G. R., I. Misztal, and R. D. Van Vleck. 1988. Animal model evaluation of Ayrshire milk yield with all lactations, herd-sire interaction, and groups based on unknown parents. J. Dairy Sci. 71:1319.

99. Wiggans, G. R., and P. M. VanRaden, 1989. USDA-DHIA animal model genetic evaluation. National Cooperative Dairy Herd Improvement Fact Sheet. H-2.

100. Wilmink, J. B. M., and J. Dommerholt. 1985. Approximate reliability of best linear unbiased prediction in models with and without relationship. J. Dairy Sci. 68:946.

APPENDIX

```
C********* An Animal Model Program *************
C by iterative Conjugate Gradient Method for a Sparse Linear System
C Input data:
C    Parents and sire ID, and sire true breeding value(4)
C    Parents and cow ID, records and true breeding value(5)
C= = = = = = = = FOR PARAMETERS = = = = = = = = = = = =
C    7M = Exact Number of Animal (= NSIRE + NCOW)
C    NSIRE = Exact Number of Sire
C    IL = Approximate Number of Fixed Effects (provide enough)
C*****************************************************************
C  Provide enough space in COMMON W and IW in main and
C  subroutine programs for nonzero elements of the coefficient matrix.
C*****************************************************************
      PARAMETER(IM = 12000,NSIRE = 500,IL = 2000)
      COMMON W(200000),IW(2,200000)
      INTEGER ILA(IM),ILPA(2,IM),ION(IM),ISMN(IM),IGC(IM),
     1      IHYS(2,IL),IXZP(2,6*IM),ITC(2,NSIRE*4),ITD(NSIRE*4),
     2      IZX(2,6*IM),IXX(2,IL),IXZ(2,6*IM),IRC(2,6*IM),IYC(IM),
     3      ILAT(IM),M(IM)
      REAL X(IL + IM),AI(6*IM),ZX(6*IM),XZ(6*IM),XX(IL),ZZ(IM),
     1    XY(IL),ZY(IM),WY(IL + IM),ZMY(IM),XXXY(IL),BETA(IL),
     2    BV(IM),RT(IM),RE(IM),TAI(NSIRE*4),REL(4,IM),
     3    BKX(6),AKY(6),BKY(6),KX(6),KY(6),AKX(6),
     4    AKZ(6,IM),BKZ(6,IM),KZ(6,IM),
     5    RTS(NSIRE),RES(NSIRE),XZP(6*IM)
      REAL*8 MSE,TEMP(IM),U(IM),P(IM)
      EXTERNAL ASUB,ATSUB
C ********* OPTIONS ****************************************
C  IGBASE: the year for genetic base (e.g.,year 1 = 1)
C  IYEAR: designated year for cows (e.g., > year 1)
C  HS = heritability of the trait
C  RP = repeatability of the trait
C  IMIN = minimum group size of management groups (e.g., 5)
C*****************************************************************
      IGBASE = 1
      IYEAR = 5
      HS = .25
      RP = .50
      IMIN = 5
C************************************************
      IPE = 1
      PV = 1.
      VA = HS
      VP = RP*PV-VA
      VE = PV-VP-VA
      IF(IPE .EQ. 1) THEN
        PALPHA = VE/VP
        UALPHA = VE/VA
      ELSE
        UALPHA = (VE + VP)/VA
```

```
      ENDIF
      IGCK = 0
      IYCK = 0
      IN = 0
      IAN = 0
      IHYP = 0
      IDF = 0
      INB = 1
      IVEC = IM
      ITEC = 0
      ITDN = 0
      NNSIRE = NSIRE
C ** Read parents ID, sire ID and true breeding value.
   3 READ(4,1011,END = 12)ISB,IDB,IBULL,TBV
      IN = IN + 1
      ILA(IN) = IBULL
      BV(IN) = TBV
      IS = IFIND(ISB,ILA,IN)
      ILPA(1,IN) = IS
      ILPA(2,IN) = IDB
C** Storing inverse of relationship matrix for sires in AI vector
      IF(ISB .EQ. 0 .AND. IDB .EQ. 0) THEN
         IRC(1,IN) = IN
         IRC(2,IN) = IN
        AI(IN) = AI(IN) + 1.
      ELSE
       IF(IDB .EQ. 0) THEN
        ION(IS) = ION(IS) + 1
        AI(IS) = AI(IS) + .33333333333
         IVEC = IVEC + 1
         IRC(1,IVEC) = IS
         IRC(2,IVEC) = IN
        AI(IVEC) = AI(IVEC) - .6666666666
         IRC(1,IN) = IN
         IRC(2,IN) = IN
        AI(IN) = AI(IN) + 1.33333333333
       GOTO 3
       ENDIF
       IF(ISB .EQ. 0) THEN
        ITDN = ITDN + 1
         ITD(ITDN) = IDB
         ITEC = ITEC + 1
         ITC(1,ITEC) = ITDN
         ITC(2,ITEC) = ITDN
        TAI(ITDN) = TAI(ITDN) + .333333333
         ITEC = ITEC + 1
         ITC(1,ITEC) = IN
         ITC(2,ITEC) = ITDN
        TAI(ITEC) = TAI(ITEC) - .666666666
         IRC(1,IN) = IN
         IRC(2,IN) = IN
        AI(IN) = AI(IN) + 1.333333333
       GOTO 3
       ENDIF
       ION(IS) = ION(IS) + 1
       ITDN = ITDN + 1
       ITD(ITDN) = IDB
```

```
      AI(IS) = AI(IS) + .5
        ITEC = ITEC + 1
        ITC(1,ITEC) = ITDN
        ITC(2,ITEC) = ITDN
       TAI(ITEC) = TAI(ITEC) + .5
        ITEC = ITEC + 1
        ITC(1,ITEC) = IS
        ITC(2,ITEC) = ITDN
       TAI(ITEC) = TAI(ITEC) + .5
        IVEC = IVEC + 1
        IRC(1,IVEC) = IS
        IRC(2,IVEC) = IN
       AI(IVEC) = AI(IVEC)-1.
        ITEC = ITEC + 1
        ITC(1,ITEC) = IN
        ITC(2,ITEC) = ITDN
       TAI(ITEC) = TAI(ITEC)-1.
        IRC(1,IN) = IN
        IRC(2,IN) = IN
       AI(IN) = AI(IN) + 2.
      ENDIF
      GOTO 3
   12 CONTINUE
      IDCP = IN
      CALL RANK(BV,NSIRE,RTS)
C** Read data (herd-year, season, parents and cow ID, and records)
    1 READ(5,1001,END = 11) IHY,ISM,IDS, IDD, IDC,Y,ING,TB
      IAN = IAN + 1
      IY = MOD(IHY,100)
      IF(IDS .EQ. 0) GOTO 1
      IN = IFIND(IDC,ILA,IDCP)
      IF(IN .EQ. 0) THEN
        IN = IDCP + 1
        REL(1,IN) = REL(1,IN) + 1
        IF(IGBASE .EQ. IY) THEN
          IGCK = IGCK + 1
          IGC(IGCK) = IN
        ENDIF
        IF(IYEAR .EQ. IY) THEN
          IYCK = IYCK + 1
          IYC(IYCK) = IN
        ENDIF
      ELSE
        REL(1,IN) = REL(1,IN) + 1
        IF(IGBASE .EQ. IY) THEN
          IGCK = IGCK + 1
          IGC(IGCK) = IN
        ENDIF
        IF(IYEAR .EQ. IY) THEN
          IYCK = IYCK + 1
          IYC(IYCK) = IN
        ENDIF
        GOTO 22
      ENDIF
      IDCP = IN
      BV(IN) = TB
      ILA(IN) = IDC
```

```
      ISMN(IN) = ISM
      ION(IN) = 0
      IS = IFIND(IDS,ILA,IN)
      ID = IFIND(IDD,ILA,IN)
      ILPA(1,IN) = IS
      ILPA(2,IN) = ID
C** Storing inverse of relationship matrix for cows in AI vector.
      IF(IS .EQ. 0 .AND. ID .EQ. 0) THEN
        IRC(1,IN) = IN
        IRC(2,IN) = IN
       AI(IN) = AI(IN) + 1.
      ELSE
       IF(ID .EQ. 0)THEN
        ION(IS) = ION(IS) + 1
        AI(IS) = AI(IS) + .33333333
          IVEC = IVEC + 1
          IRC(1,IVEC) = IS
          IRC(2,IVEC) = IN
        AI(IVEC) = AI(IVEC) - .666666666
          IRC(1,IN) = IN
          IRC(2,IN) = IN
        AI(IN) = AI(IN) + 1.333333333
       GOTO 22
       ENDIF
       IF(IS .EQ. 0) THEN
        ION(ID) = ION(ID) + 1
        AI(ID) = AI(ID) + .333333333
          IVEC = IVEC + 1
          IRC(1,IVEC) = ID
          IRC(2,IVEC) = IN
        AI(IVEC) = AI(IVEC) - .666666666
          IRC(1,IN) = IN
          IRC(2,IN) = IN
        AI(IN) = AI(IN) + 1.3333333333
        GOTO 22
       ENDIF
       ION(IS) = ION(IS) + 1
       ION(ID) = ION(ID) + 1
       AI(IS) = AI(IS) + .5
       AI(ID) = AI(ID) + .5
         CALL OFFD(IVECD,IRC,IVEC,IS,ID)
         IRC(1,IVECD) = IS
         IRC(2,IVECD) = ID
       AI(IVECD) = AI(IVECD) + .5
         IVEC = IVEC + 1
         IRC(1,IVEC) = IS
         IRC(2,IVEC) = IN
       AI(IVEC) = AI(IVEC)-1.
         IVEC = IVEC + 1
         IRC(1,IVEC) = ID
         IRC(2,IVEC) = IN
       AI(IVEC) = AI(IVEC)-1.
         IRC(1,IN) = IN
         IRC(2,IN) = IN
       AI(IN) = AI(IN) + 2.
      ENDIF
      GOTO 22
```

```
  11 IEND = 1
C* Storing relationship matrix for dams of bulls in AI vector.
     ITVEC = IVEC
     DO 26 I = 1,ITEC
       DO 29 J = NSIRE + 1,IVEC
     IF(ITD(ITC(2,I)) .EQ. ILA(IRC(2,J)) .AND.
      *ITD(ITC(1,I)) .EQ. ILA(IRC(1,J)))THEN
          AI(J) = AI(J) + TAI(I)
          GOTO 26
        ENDIF
  29     CONTINUE
     ITTD = ITD(ITC(2,I))
     ID = IFIND(ITTD,ILA,IN)
     ITVEC = ITVEC + 1
     IRC(1,ITVEC) = ITC(1,I)
     IRC(2,ITVEC) = ID
     AI(ITVEC) = TAI(I)
  26   CONTINUE
     IVEC = ITVEC
     DO 30 I = 1,IVEC
       IF(IRC(1,I) .NE. IRC(2,I)) THEN
         ITVEC = ITVEC + 1
         IRC(1,ITVEC) = IRC(2,I)
         IRC(2,ITVEC) = IRC(1,I)
         AI(ITVEC) = AI(I)
       ENDIF
  30   CONTINUE
     IVEC = ITVEC
C** Coefficient Matrix and Right Hand Side **
     GOTO 23
  22 CONTINUE
     IF(IHYP .EQ. IHY .OR. IAN .EQ. 1) THEN
       IF(IDS .NE. 0) THEN
         IF(IN .GT. INP)INP = IN
         IF(IN .LT. INB)INB = IN
       ENDIF
       GOTO 24
     ENDIF
C** Combining herd-year-seasons to form management groups
  23 IBC = 0
     IMP = 0
     IAL = 1
     DO 96 I = 1,6
       IF(KX(I) .NE. 0) THEN
         IPM = IPM + 1
         SKX = SKX + KX(I)*KX(I)
         SMX = SMX + KX(I)
       ENDIF
       BKX(I) = 0.
       BKY(I) = 0.
  96   CONTINUE
     DO 47 I = 1,6
       IF(KX(I) .EQ. 0) GOTO 47
       IP = I + 1
       IA = 1
       AKX(IA) = KX(I)
       AKY(IA) = KY(I)
```

```
      DO 95 K = INB,INP
95      AKZ(IA,K) = KZ(I,K)
    IAF = IAF-KX(I)
    IF(KX(I) .GE. IMIN .AND. IAF .GE. IMIN) GOTO 97
    DO 48 J = IP,6
      IF(KX(J) .EQ. 0) GOTO 48
      IDF = IDF + 1
      IA = IA + 1
      KX(I) = KX(J) + KX(I)
      KY(I) = KY(J) + KY(I)
      IAF = IAF-KX(J)
      AKX(IA) = KX(J)
      AKY(IA) = KY(J)
      KX(J) = 0.
      KY(J) = 0.
      DO 45 K = INB,INP
45        KZ(I,K) = KZ(J,K) + KZ(I,K)
      IF(KX(I) .GE. IMIN .AND. IAF .GE. IMIN) GOTO 97
48    CONTINUE
97      IMI = I
    IF(IMP .EQ. 0) GOTO 79
    IF(KX(IMP) .EQ. KX(IMI)) GOTO 46
    IF(KX(IMP) .LT. KX(IMI)) GOTO 150
    IF(IAL .EQ. IAP) GOTO 44
    DO 31 J = IAL,IAP
      JP = IAP-J + 1
      DIF = ABS(KX(IMI)-KX(IMP))
      MAD = KX(IMP)-BKX(JP)
      MSD = KX(IMI) + BKX(JP)
      DIFP = ABS(MSD-MAD)
      IF(DIF .LE. DIFP) GOTO 44
      KX(IMP) = MAD
      KX(IMI) = MSD
      KY(IMP) = KY(IMP)-BKY(JP)
      KY(IMI) = KY(IMI) + BKY(JP)
      DO 31 K = INB,INP
        KZ(IMP,K) = KZ(IMP,K)-BKZ(JP,K)
31        KZ(IMI,K) = KZ(IMI,K) + BKZ(JP,K)
150       CONTINUE
    DO 93 J = 1,IA
      DIF = ABS(KX(IMI)-KX(IMP))
      MAD = KX(IMP) + AKX(J)
      MSD = KX(IMI)-AKX(J)
      DIFP = ABS(MSD-MAD)
      IF(DIF .LE. DIFP) GOTO 79
      IAL = J
      IF(IA .EQ. 1) GOTO 79
      KX(IMP) = MAD
      KX(IMI) = MSD
      KY(IMP) = KY(IMP) + AKY(J)
      KY(IMI) = KY(IMI)-AKY(J)
      DO 93 K = INB,INP
        KZ(IMP,K) = KZ(IMP,K) + AKZ(J,K)
93        KZ(IMI,K) = KZ(IMI,K)-AKZ(J,K)
44    IAL = 1
79    CONTINUE
    DO 21 IC = IAL,IA
```

```fortran
      BKX(IC) = AKX(IC)
      BKY(IC) = AKY(IC)
      DO 21 K = INB,INP
 21      BKZ(IC,K) = AKZ(IC,K)
      IAP = IA
 46   CONTINUE
      IBC = IBC + 1
      IF(IBC .EQ. 1) THEN
        IMP = I
      ELSE
        IMP = IMI
      ENDIF
 47 CONTINUE
      DO 27 I = 1,6
      IF(KX(I) .EQ. 0) GOTO 27
      IHN = IHN + 1
      IHYS(1,IHN) = IHYP
      IHYS(2,IHN) = I
        IXX(1,IHN) = IHN
        IXX(2,IHN) = IHN
      XX(IHN) = KX(I)
      WY(IHN) = KY(I)
      DO 28 J = INB,INP
        IF(KZ(I,J) .EQ. 0.) GOTO 28
        KV = KV + 1
        IXZ(1,KV) = IHN
        IXZ(2,KV) = J
        XZ(KV) = KZ(I,J)
        LV = LV + 1
        IZX(1,LV) = J
        IZX(2,LV) = IHN
        ZX(LV) = KZ(I,J)
 28      CONTINUE
 27 CONTINUE
      DO 25 I = 1,6
      KY(I) = 0.
      KX(I) = 0.
      DO 25 J = INB,INP
        KZ(I,J) = 0.
 25   CONTINUE
      IF(IEND .EQ. 1) GOTO 58
      INP = IN
      INB = IN
      IAF = 0.
 24 IHYP = IHY
      KX(ISM) = KX(ISM) + 1.
      KY(ISM) = KY(ISM) + Y
      IAF = IAF + 1
      IF(IDS .EQ. 0) GOTO 1
      KZ(ISM,IN) = 1.
      ZZ(IN) = ZZ(IN) + 1.
      ZY(IN) = ZY(IN) + Y
      GOTO 1
 58 IHNP = IHN
      IN = IDCP
      NCOW = IN-NSIRE
C** THE END OF READING DATA
```

```
      ILC = IHN + IN
      DO 74 I = 1,IN
      INN = IHN + I
74    WY(INN) = ZY(I)
      CALL SORT(IVEC,IRC,AI,1,1)
      IF(IPE .EQ. 1) THEN
       VR = UALPHA/PALPHA
       MV = 0
       IXZC = IXZ(1,1)
       DO 55 I = 1,KV
        IF(IXZ(1,I) .NE. IXZC .OR. I .EQ. KV) THEN
          I1 = 1
          DO 52 J = 1,IN
            IF(TEMP(J) .EQ. 0) GOTO 52
            MV = MV + 1
            IXZP(1,MV) = I
            IXZP(2,MV) = J
            XZP(MV) = TEMP(J)
            TEMP(J) = 0.
52          CONTINUE
        ENDIF
        IXZC = IXZ(1,I)
        DO 51 J = I1,IVEC
          I1 = J
          IF(IXZ(2,I) - IRC(1,J)) 51,54,55
54        CONTINUE
          IF(IRC(1,J) .EQ. IRC(2,J)) THEN
            TEMP(IRC(2,J)) = TEMP(IRC(2,J)) + XZ(I)*(1. + AI(J)*VR)
          ELSE
            TEMP(IRC(2,J)) = TEMP(IRC(2,J)) + XZ(I)*AI(J)*VR
          ENDIF
51        CONTINUE
55      CONTINUE
        DO 53 J = 1,MV
          IXZ(1,J) = IXZP(1,J)
          IXZ(2,J) = IXZP(2,J)
          XZ(J) = XZP(J)
53      CONTINUE
       KV = MV
      ENDIF
      I3 = 1
      SSQX = .0
      DO 56 I = 1,IHN
       SSQX = SSQX + XX(I)*XX(I)
       SX = SX + XX(I)
       LVEC = LVEC + 1
       IW(1,LVEC) = I
       IW(2,LVEC) = I
       W(LVEC) = XX(I)
       DO 57 J = I3,KV
         I3 = J
         IF(I .NE. IXZ(1,J)) GOTO 56
         LVEC = LVEC + 1
         IW(1,LVEC) = IXZ(1,J)
         IW(2,LVEC) = IXZ(2,J) + IHN
         W(LVEC) = XZ(J)
57       CONTINUE
```

```
   56   CONTINUE
      SD = SQRT((SKX- SMX*SMX/IPM)/(IPM-1))
      SMEAN = SMX/IPM
      WRITE(6,1022)IPM
      WRITE(6,1021) SMEAN,SD
      SD = SQRT((SSQX- SX*SX/IHN)/(IHN-1))
      SMEAN = SX/IHN
C ** Number of herd-year-seasons, records/h-y-s **
      WRITE(6,1020) IHN
      WRITE(6,1019) SMEAN,SD
      CALL SORT(LV,IZX,ZX,1,1)
      IU = 0
      I5 = 1
      I6 = 1
      DO 61 I = 1,IN
        DO 62 J = I5,LV
          IF(I .NE. IZX(1,J)) GOTO 64
          LVEC = LVEC + 1
          IW(1,LVEC) = IZX(1,J) + IHN
          IW(2,LVEC) = IZX(2,J)
          W(LVEC) = ZX(J)
   62     CONTINUE
   64     I5 = J
          DO 63 J = I6,IVEC
            I6 = J
            IF(I .NE. IRC(1,J)) GOTO 68
            IF(IPE .EQ. 1) THEN
              LVEC = LVEC + 1
              IW(1,LVEC) = IRC(1,J) + IHN
              IW(2,LVEC) = IRC(2,J) + IHN
              IF(IRC(1,J) .EQ. IRC(2,J)) THEN
                W(LVEC) = (1. + AI(J)*VR)*ZZ(IRC(1,J)) + AI(J)*UALPHA
              ELSE
                W(LVEC) = AI(J)*VR*ZZ(IRC(1,J)) + AI(J)*UALPHA
              ENDIF
            ELSE
              LVEC = LVEC + 1
              IW(1,LVEC) = IRC(1,J) + IHN
              IW(2,LVEC) = IRC(2,J) + IHN
              W(LVEC) = AI(J)*UALPHA
              IF(IRC(1,J) .EQ. IRC(2,J)) THEN
                IU = IU + 1
                W(LVEC) = ZZ(IU) + W(LVEC)
              ENDIF
            ENDIF
   63     CONTINUE
   68     CONTINUE
   61   CONTINUE
C ** Call the subroutine for a solution vector.
      DO 38 I = 1,ILC
        X(I) = 0.0
   38   CONTINUE
      CALL SPARSE(WY,ILC,ASUB,X,RSQ,LVEC,ATSUB)
C** Calculation of Reliability ****
      KD = (4-2*HS)/HS
      DO 91 I = NSIRE + 1,IN
C** Contribution to sire from daughters' records
```

```fortran
      IF(ILPA(1,I) .NE. 0) THEN
        DN = .39 + .61/REL(1,I)
        DEDAU = 1/(.16 + .84/DN)
        REL(3,I) = DEDAU
        REL(2,ILPA(1,I)) = REL(2,ILPA(1,I)) + DEDAU
        REL(1,ILPA(1,I)) = REL(2,ILPA(1,I))/(REL(2,ILPA(1,I)) + KD)
      ENDIF
C** Reliability from her own records
      REL(1,I) = REL(1,I)*HS/(1 + (REL(1,I)-1)*RP)
      REL(2,I) = KD*REL(1,I)/(1-REL(1,I))
 91   CONTINUE
      DO 92 I = IN,NNSIRE + 1,-1
        IF(ILPA(2,I) .NE. 0) THEN
          RELSIRE = REL(2,ILPA(1,I))-1/(.16 + .84/REL(2,I))
          RELSIRE = RELSIRE/(RELSIRE + KD)
          RELANIM = REL(1,I)/(4-REL(1,I)*RELSIRE)
          DEDAU = KD*RELANIM/(1-RELANIM)
          REL(4,I) = DEDAU
          REL(2,ILPA(2,I)) = REL(2,ILPA(2,I)) + DEDAU
          REL(1,ILPA(2,I)) = REL(2,ILPA(2,I))/(REL(2,ILPA(2,I)) + KD)
        ENDIF
 92   CONTINUE
      DO 98 I = NNSIRE,1,-1
        ITDB = ILPA(2,I)
        ID = IFIND(ITDB,ILA,IN)
        ILPA(2,I) = ID
        IF(ILPA(1,I) .NE. 0) THEN
          IF(ILPA(2,I) .NE. 0) THEN
            RELDAM = REL(1,ILPA(2,I))
          ELSE
            RELDAM = 0
          ENDIF
          RELANIM = REL(1,I)/(4-REL(1,I)*RELDAM)
          DESON = KD*RELANIM/(1-RELANIM)
          REL(3,I) = DESON
          REL(2,ILPA(1,I)) = REL(2,ILPA(1,I)) + DESON
          REL(1,ILPA(1,I)) = REL(2,ILPA(1,I))/(REL(2,ILPA(1,I)) + KD)
          DESON = .0
        ENDIF
        IF(ILPA(2,I) .NE. 0) THEN
          IF(ILPA(1,I) .NE. 0) THEN
            RELSIRE = REL(1,ILPA(1,I))
          ELSE
            RELSIRE = 0
          ENDIF
          RELANIM = REL(1,I)/(4-REL(1,I)*RELSIRE)
          DESON = KD*RELANIM/(1-RELANIM)
          REL(4,I) = DESON
          REL(2,ILPA(2,I)) = REL(2,ILPA(2,I)) + DESON
          REL(1,ILPA(2,I)) = REL(2,ILPA(2,I))/(REL(2,ILPA(2,I)) + KD)
          DESON = .0
        ENDIF
 98   CONTINUE
      DO 94 I = 1,IN
        IF(ILPA(1,I) .NE. 0) THEN
          DE = REL(2,ILPA(1,I))-REL(3,I)
          RELSIRE = DE/(DE + KD)
```

```
      ELSE
        RELSIRE = 0
      ENDIF
      IF(ILPA(2,I) .NE. 0) THEN
        DE = REL(2,ILPA(2,I))-REL(4,I)
        RELDAM = DE/(DE + KD)
      ELSE
        RELDAM = 0
      ENDIF
        RELPA = (RELSIRE + RELDAM)/4
        DEPA = KD*RELPA/(1-RELPA)
        REL(2,I) = DEPA + REL(2,I)
        REL(1,I) = REL(2,I)/(REL(2,I) + KD)
  94  CONTINUE
C*******************************************
      DO 84 K = 1,IL
  84  BETA(K) = X(K)
      DO 86 K = 1,IM
        I2 = K + IHN
        U(K) = X(I2)
  86  CONTINUE
      IF(IPE .EQ. 1) THEN
        DO 81 I = 1,IVEC
          P(IRC(1,I)) = P(IRC(1,I)) + AI(I)*VR*U(IRC(2,I))
  81    CONTINUE
      ENDIF
      WRITE(6,1002)LVEC
      WRITE(6,1012)
      DO 89 I = 1,IHNP
        WRITE(6,1013)(IHYS(J,I),J = 1,2),BETA(I),XX(I)
  89  CONTINUE
  80 CONTINUE
      WRITE(6,1009)IYEAR
      DPP = .0
C ** Output for all animals
C     DO 59 I = 1,IN
C       IF(I .LE. NSIRE) THEN
C       WRITE(6,1000)(ILPA(J,I),J = 1,2),ILA(I),U(I),DPP,ION(I),IHYN(I),
C   1             ISMN(I)
C       ELSE
C       WRITE(6,1000)(ILPA(J,I),J = 1,2),ILA(I),U(I),P(I),ION(I),IHYN(I),
C   1             ISMN(I)
C       ENDIF
C 59    CONTINUE
      GBASE = .0
      DO 69 I = 1,IGCK
        GBASE = GBASE + U(IGC(I))/2
  69  CONTINUE
      GBASE = GBASE/IGCK
      DPP = 0.
c ** Output for a certain year.
      DO 67 I = 1,NSIRE
        U(I) = U(I)/2-GBASE
       WRITE(6,1000)ILA(ILPA(1,I)),ILA(ILPA(2,I)),ILA(I),U(I),DPP,
      *           ION(I),REL(1,I)
  67  CONTINUE
      WRITE(6,1009)IYEAR
```

```
      DO 70 I = 1,IYCK
        P(IYC(I)) = U(IYC(I)) + P(IYC(I))-2*GBASE
        U(IYC(I)) = U(IYC(I))/2-GBASE
        ILAT(I) = ILA(IYC(I))
   70   CONTINUE
      CALL SORT1(ILAT,IYCK,M)
      DO 71 I = 1,IYCK
        K = M(I)
        WRITE(6,1000)ILA(ILPA(1,IYC(K))),ILA(ILPA(2,IYC(K))),
   1       ILA(IYC(K)),U(IYC(K)),
   1       P(IYC(K)),ION(IYC(K)),REL(1,IYC(K))
   71   CONTINUE
      DO 72 I = 1,NSIRE
        ER = BV(I)-U(I)
        MSE = MSE + ER*ER
        SET = SET + BV(I)*U(I)
        SE = SE + U(I)
        ST = ST + BV(I)
        SSE = SSE + U(I)*U(I)
        SST = SST + BV(I)*BV(I)
   72   CONTINUE
      CVCO = (SET-(SE*ST)/NSIRE)
      EVCO = (SSE-(SE*SE)/NSIRE)
      TVCO = (SST-(ST*ST)/NSIRE)
      BVCO = CVCO/SQRT(EVCO*TVCO)
      WRITE(6,1015)BVCO
      SET = .0
      SST = .0
      SSE = .0
      SE = .0
      ST = .0
      DO 75 I = NSIRE + 1, IN
        ER = BV(I)-U(I)
        MSE = MSE + ER*ER
        SET = SET + BV(I)*U(I)
        SE = SE + U(I)
        ST = ST + BV(I)
        SSE = SSE + U(I)*U(I)
        SST = SST + BV(I)*BV(I)
   75   CONTINUE
      CVCO = (SET-(SE*ST)/NCOW)
      EVCO = (SSE-(SE*SE)/NCOW)
      TVCO = (SST-(ST*ST)/NCOW)
      BVCO = CVCO/SQRT(EVCO*TVCO)
      WRITE(6,1016)BVCO
      DO 78 I = 1,IN
        SK = SK + 1.
        DSK = DSK + SK*SK
   78   CONTINUE
      PEV = PEV*VE/IN
      MSE = MSE/IN
      CVM = (CTE-(SEM*STM)/NCOW)
      EVM = (SSEM-(SEM*SEM)/NCOW)
      TVM = (SSTM-(STM*STM)/NCOW)
      WRITE(6,1014)MSE
 1000 FORMAT(3I8,2F12.2,I5,F8.3)
 1001 FORMAT(I5,I1,I7,I6,I6,F13.2,I7,F10.2)
```

```
1002 FORMAT(' NO. OF NONZERO ELEMENTS IN COEFF. MATRIX', I8)
1003 FORMAT(13F7.2)
1004 FORMAT(3I8,2F12.2,I5,I7,I1,F8.0,F8.0)
1009 FORMAT('  SIRE    DAM  ANIMAL      PTA        PPA',
   1    ' NO. PROGENY  IN YEAR ', I2)
1011 FORMAT(3I6, F10.2)
1012 FORMAT(' HERD-YEAR-SEASON EFFECT   NO. OF G. MATES')
1013 FORMAT(I6,I1, F15.2,F15.0)
1014 FORMAT(' MSE = ',F12.2 )
1015 FORMAT(' SIRE  B.V. CORRELATION = ',F6.3)
1016 FORMAT(' COW  B.V. CORRELATION = ',F6.3)
1019 FORMAT(' AVERAGE SIZE OF M. GROUPS = ',F7.2,'   STD =', F7.2)
1020 FORMAT(' NO. OF MANAGEMENT GROUPS = ' ,I5)
1021 FORMAT(' AVERAGE SIZE OF HYS = ',F7.2,'   STD=', F7.2)
1022 FORMAT(' NO. OF HYS = ' ,I5)
     STOP
     END
     FUNCTION IFIND(ID,LIST,N)
     INTEGER LIST(N)
     DO 1 I=1,N
     IF(ID .EQ. LIST(I)) GOTO 2
   1 CONTINUE
     IFIND=0
     RETURN
   2 CONTINUE
     IFIND=I
     RETURN
     END
     SUBROUTINE OFFD(IVECD,IRC,IVEC,IS,ID)
     DIMENSION IRC(2,IVEC)
     IVECD=0
     DO 1 I=1,IVEC
       IF(IRC(1,I) .EQ. IS .AND. IRC(2,I) .EQ. ID ) IVECD=I
   1 CONTINUE
       IF(IVECD .GT. 0) GOTO 2
     IVECD=IVEC+1
   2 RETURN
     END
     SUBROUTINE SORT(IVEC,IRC,AI,IB,IC)
C**************************************************************
C Option 1
C    IB=1 sorting rows and columns
C    IC=1 sorting by row first,  IC=2 by column first.
C **************************************************************
C Option 2
C    IB=0, IC=1  sorting by row only or
C    IB=0, IC=2  sorting by column only.
C**************************************************************
     DIMENSION IRC(2,IVEC),AI(IVEC)
     IF(IC .EQ. 1) THEN
       ID=2
     ELSE
       ID=1
       IC=2
     ENDIF
     NM1=IVEC-1
     DO 32 I=1,NM1
```

```
      K = I
      IP1 = I + 1
      DO 33 J = IP1,IVEC
       IF(IRC(IC,J) .LT. IRC(IC,K))K = J
 33    CONTINUE
      DO 41 JJ = 1,2
       IG1 = IRC(JJ,I)
       IRC(JJ,I) = IRC(JJ,K)
 41    IRC(JJ,K) = IG1
      W1 = AI(I)
      AI(I) = AI(K)
      AI(K) = W1
 32   CONTINUE
      IF(IB .EQ. 0) GOTO 45
      IMP = 1
      ISP = 1
      IPP = IRC(IC,1)
 38 CONTINUE
      IPY = 0
      DO 34 L = IMP + 1,IVEC
       IF(IRC(IC,L) .NE. IPP) GOTO 35
      IPY = IPY + 1
      IPP = IRC(IC,L)
 34    CONTINUE
 35    CONTINUE
      IMP = L
      IEP = ISP + IPY
      IF(IPY .NE. 0) THEN
 37   NM1 = IEP-1
      DO 43 I = ISP,NM1
       K = I
       IP1 = I + 1
       DO 52 J = IP1,IEP
        IF(IRC(ID,J) .LT. IRC(ID,K))K = J
 52     CONTINUE
       DO 51 JJ = 1,2
        IG1 = IRC(JJ,I)
        IRC(JJ,I) = IRC(JJ,K)
 51     IRC(JJ,K) = IG1
       W1 = AI(I)
       AI(I) = AI(K)
       AI(K) = W1
 43    CONTINUE
      ENDIF
      ISP = IMP
      IPP = IRC(IC,ISP)
      IF(IVEC .GE. ISP) GOTO 38
 45 RETURN
      END
      SUBROUTINE SPARSE(B,N,ASUB,X,RSQ,IV,ATSUB)
      PARAMETER (NMAX = 11500,EPS = 1.E-6)
      DIMENSION B(N),X(N),G(NMAX),H(NMAX),XI(NMAX),XJ(NMAX)
      EPS2 = N*EPS**2
      IRST = 0
 1    IRST = IRST + 1
      CALL ASUB(X,XI,N,IV)
      RP = 0.
```

```
      BSQ = 0.
      DO 11 J = 1,N
        BSQ = BSQ + B(J)**2
        XI(J) = XI(J)-B(J)
        RP = RP + XI(J)**2
11    CONTINUE
      CALL ATSUB(XI,G,N,IV)
      DO 12 J = 1,N
        G(J) = -G(J)
        H(J) = G(J)
12    CONTINUE
      DO 19 ITER = 1,10*N
        CALL ASUB(H,XI,N,IV)
        ANUM = 0.
        ADEN = 0.
        DO 13 J = 1,N
          ANUM = ANUM + G(J)*H(J)
          ADEN = ADEN + XI(J)**2
13      CONTINUE
        IF(ADEN.EQ.0.)PAUSE 'very singular matrix'
        ANUM = ANUM/ADEN
        DO 14 J = 1,N
          XI(J) = X(J)
          X(J) = X(J) + ANUM*H(J)
14      CONTINUE
        CALL ASUB(X,XJ,N,IV)
        RSQ = 0.
        DO 15 J = 1,N
          XJ(J) = XJ(J)-B(J)
          RSQ = RSQ + XJ(J)**2
15      CONTINUE
        IF(RSQ.EQ.RP.OR.RSQ.LE.BSQ*EPS2)RETURN
        IF(RSQ.GT.RP)THEN
          DO 16 J = 1,N
            X(J) = XI(J)
16        CONTINUE
          IF(IRST.GE.3)RETURN
          GO TO 1
        ENDIF
        RP = RSQ
        CALL ATSUB(XJ,XI,N,IV)
        GG = 0.
        DGG = 0.
        DO 17 J = 1,N
          GG = GG + G(J)**2
          DGG = DGG + (XI(J) + G(J))*XI(J)
17      CONTINUE
        IF(GG.EQ.0.)RETURN
        GAM = DGG/GG
        DO 18 J = 1,N
          G(J) = -XI(J)
          H(J) = G(J) + GAM*H(J)
18      CONTINUE
19    CONTINUE
      PAUSE 'too many iterations'
      RETURN
      END
```
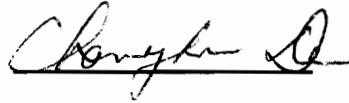
```
      SUBROUTINE ASUB(X,V,N,IV)
      DIMENSION X(N),V(N)
      COMMON W(200000),IW(2,200000)
      DO 12 I=1,N
 12   V(I)=0.
      DO 3 J=1,IV
         V(IW(1,J))=V(IW(1,J))+X(IW(2,J))*W(J)
  3      CONTINUE
      RETURN
      END
      SUBROUTINE ATSUB(X,V,N,IV)
      DIMENSION X(N),V(N)
      COMMON W(200000),IW(2,200000)
      DO 12 I=1,N
 12   V(I)=0.
      DO 3 J=1,IV
         V(IW(2,J))=V(IW(2,J))+X(IW(1,J))*W(J)
  3      CONTINUE
      RETURN
      END
      SUBROUTINE SORT1(X,N,M)
      DIMENSION X(1),M(1)
      DO 200 I=1,N
      K=0.
      DO 100 J=1,N
       IF(X(I) .GT. X(J) .OR. (X(I) .EQ. X(J) .AND. I .GE. J))K=K+1
100     CONTINUE
      M(K)=I
200   CONTINUE
      RETURN
      END
```

VITA

The author was born on January 4, 1957 in Daejeon, Korea and was raised in the farm family until finished Shindoh elementary school. The author enrolled at Chungnam National University in 1976, served in the army for 3 years, and completed summer internships at National Institute of Seedstock in Seong-hwan and National Livestock Experimentation Station in Suwon, before graduating with a B.S. in Dairy Science in 1983. The author obtained the M.S. degree in Animal Science at Seoul National University in 1985 and completed the requirements for the Ph.D. in Animal Science-Dairy Breeding at Virginia Polytechnic Institute and State University in 1992.

Changhee Do