

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Faculty Papers and Publications in Animal
Science

Animal Science Department

1-1-1983

Genetic Groups in Dairy Sire Evaluation Under a Selection Model

T. R. Famula
Cornell University

E. J. Pollak
Cornell University, jpollak2@unl.edu

L. Dale Van Vleck
University of Nebraska-Lincoln, dvan-vleck1@unl.edu

Follow this and additional works at: <https://digitalcommons.unl.edu/animalscifacpub>



Part of the [Animal Sciences Commons](#)

Famula, T. R.; Pollak, E. J.; and Van Vleck, L. Dale, "Genetic Groups in Dairy Sire Evaluation Under a Selection Model" (1983). *Faculty Papers and Publications in Animal Science*. 392.
<https://digitalcommons.unl.edu/animalscifacpub/392>

This Article is brought to you for free and open access by the Animal Science Department at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Faculty Papers and Publications in Animal Science by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

Genetic Groups in Dairy Sire Evaluation Under a Selection Model

T. R. FAMULA,¹ E. J. POLLAK, and L. D. VAN VLECK

Animal Science Department
Cornell University
Ithaca, NY 14853

ABSTRACT

The frequently arbitrary application and definition of genetic groups makes sire grouping the "weak link" of the sire evaluation process. A selection model for records not in the model for observations is similar in form to the genetic groups model. Implications of this correspondence are discussed and illustrated through a small example. Guidelines for possible grouping strategies also are discussed.

INTRODUCTION

Notwithstanding the many research contributions on application of linear model theory to animal breeding, comparison of alternative models remains a perplexing problem. Because construction of linear models often requires untestable assumptions, disagreement over proper formulation of the model inevitably follows. This is the basis of one of the current problems facing animal breeders. The disagreement concerns arbitrary grouping of sires in linear models for genetic evaluation. Questions include: 1) are genetic groups for sires necessary, and 2) if necessary, how shall they be defined? This paper examines the role of selection in the use of genetic groups and offers guidelines for possible grouping strategies.

Genetic groups are used to represent populations of sires for which the mean breeding value may differ from population to population. The word population is not intended here to describe a community of potentially interbreeding individuals; rather a population may be defined as all individuals born in the same year or

geographic area. Factors that encourage genetic groups, so defined, result from the success of current selection programs for sires and dams and the widespread use of frozen semen. The increased rate of genetic progress has complicated comparison of old and young sires.

Yet the possibilities for grouping strategies are endless, and for this reason, many animal breeders question arbitrary distinctions often made between sire groups. Failure to describe adequately the structure of genetic groups makes sire grouping the "weak link" of sire evaluation. Although a concise, unique algebraic representation for genetic groups is desirable, this goal appears unattainable.

A MIXED LINEAR MODEL WITH GENETIC GROUPS

Consider the mixed linear model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e} \quad [1]$$

where \mathbf{y} is a vector of observations of length n , \mathbf{X} is a known $n \times p$ incidence matrix, $\boldsymbol{\beta}$ is an unknown vector of fixed effects, \mathbf{z} is a known $n \times t$ incidence matrix, $\mathbf{u} = \mathbf{Q}\mathbf{g} + \mathbf{s}$ for \mathbf{Q} a known $t \times r$ incidence matrix, and \mathbf{g} a vector of unknown fixed effects. The \mathbf{s} and \mathbf{e} are unobservable random vectors with null means and accordingly,

$$E[\mathbf{u}] = \mathbf{Q}\mathbf{g}$$

$$E[\mathbf{y}] = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{Q}\mathbf{g}$$

and

$$\text{Var} \begin{bmatrix} \mathbf{s} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbf{G} & \mathbf{O} \\ \mathbf{O} & \mathbf{R} \end{bmatrix} \sigma^2$$

for some, possible unknown, scalar σ^2 . Both \mathbf{G} and \mathbf{R} are known, nonsingular covariance matrices. An equivalent form of [1] is

Received March 12, 1982.

¹Animal Science Department, University of California, Davis 95616.

$$y = X\beta + ZQg + Zs + e \quad [2]$$

For the problems addressed in this paper, Q is an incidence matrix that classifies each of t sires into r groups, g represents mean breeding values of the r groups, s is a sire's deviation from his group mean, and G is the matrix of numerator relationships among sires. The best linear unbiased predictor (BLUP), is a no selection model, of a sire's transmitting ability, or "proof" is $u = Qg + s$ where g and s may be computed from the mixed model equations:

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}ZQ & X'R^{-1}Z \\ Q'Z'R^{-1}X & Q'Z'R^{-1}ZQ & Q'Z'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}ZQ & Z'R^{-1}Z + G^{-1} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{g} \\ \hat{s} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Q'Z'R^{-1}y \\ Z'R^{-1}y \end{bmatrix} \quad [3]$$

The Qg is rarely estimable, and, therefore, $Q\hat{g}$ is not necessarily unique. Thus, comparisons of sire proofs must be based on some estimable function of g . For example, because differences between group effects are estimable, some function of group differences might be used. The "base" of the Northeast Artificial Insemination Sire Comparison (NEAISC) is analogous, although not identical, to choosing one group from which all differences between groups are taken. Because this is an estimable function of g , sire proofs are unique and invariant to the solution of [3].

SELECTION AND GENETIC GROUPS

The selection of sires and dams of sires is responsible for the differences in expected breeding value from one sire population to the next. The role of genetic groups is to account for differences in expected breeding values of sires from population to population. In this section the association of genetic groups with particular selection models is discussed.

Recall the genetic groups model of [2] and the ensuing mixed model equations of [3]. A modification of [3] that yields a direct solution for genetic proofs, i.e., $\hat{u} = Q\hat{g} + \hat{s}$, has been established by Quaas and Pollak (13). Their

derivation begins by premultiplying each side of equation [3] by

$$T = \begin{bmatrix} I & O & O \\ O & I & -Q' \\ O & O & I \end{bmatrix}$$

and inserting T' (T')⁻¹ = I between the coefficient matrix and the solution vector of [3]. Simplified, the modified equations are

$$\begin{bmatrix} X'R^{-1}X & O & X'R^{-1}Z \\ O & Q'G^{-1}O & -Q'G^{-1} \\ Z'R^{-1}X & -G^{-1}Q & Z'R^{-1}Z + G^{-1} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{g} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ O \\ Z'R^{-1}y \end{bmatrix} \quad [4]$$

As mentioned by Quaas and Pollak (13), the advantage of [4] over [3] is not necessarily computational. Instead, the modified equations provide a direct examination of BLUP sire proofs when genetic groups are used.

To examine the association between a genetic model and a particular selection model, recall the selection model presented by Henderson (7). Henderson (7) derived best linear unbiased estimators and predictors for a model conditional on selection of a random variable jointly distributed with the random variables of the usual linear model [1]. This additional random variable is called the conditional variable, and as a result of selection, the conditional variable has a different mean and variance from those the same variable would have in the model of no selection (unconditional model). Henderson's (7) example of a conditional variable is the difference between the mean of first lactation cows that go on to a second lactation and the first lactation records of cows that do not. The term selection is used in the sense that the usual assumptions of random sampling involved for estimation, prediction, and hypothesis testing have been violated.

Several assumptions are critical to the Henderson (7) selection model, and they should be presented before further discussion of genetic

groups. The first assumption is that the observation vector y is a sample from a multivariate normal distribution. This is necessary for the application of a result of Pearson (10) used to derive the mean and variance of a population after selection. Moreover, the Pearson result was derived for sequential selection of one cycle only, and its application to populations under continuous (or multicycle) selection is not justified. Intuitively, however, one might argue that because the conditional distribution of a normal is normal also, the Pearson result can be extended to more than one cycle of selection. The work of this paper, which makes use of Henderson's model, relies on the extension of this result. As with BLUP in a no selection model, the true variances must be known. Furthermore, [3] is BLUP in the absence of selection and is simultaneously BLUP in a particular selection model (7), given the conditions described.

Henderson (7) proceeds from derivation of general estimators and predictors to consider three special cases of selection: $L'y$, $L'u$, and $L'e$ for some matrix of full row rank L' . Under $L'u$ selection the mixed model equations are

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z & 0 \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} & -L \\ 0 & -L' & L'GL \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{u} \\ \hat{t} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \\ 0 \end{bmatrix} \quad [5]$$

(7). Note that [4] and [5] are identical systems of equations if $L' = Q'G^{-1}$ (or $Q = GL$) and $\hat{t} = \hat{g}$. Thus, for any $L'u$ selection model there is an analogous groups model (13).

Implications of the correspondence between the groups model of [2] and the $L'u$ selection model are not obvious. In proposing $L'u$ selection, Henderson (7) visualized a form of selection where one has "prior information" that some sires are superior to other sires. This could be the result of selection of the dams of the sires (dams not included in the evaluation) or selection on observations not included in the model. Sire groups become a means of accounting for unavailable information (or observa-

tions). However, as Quaas and Pollak (13) mention, because $Q = GL$, knowledge of the form of L can lead directly to an appropriate structure for Q . Incorporating numerator relationship coefficients into the grouping structure is analogous to the grouping definitions proposed by Henderson (6) and Thompson (17), and this result supports their suggestions.

When selection decisions that influence the choice of sires are based on information available in the mixed model equations (i.e., progeny means or previous proofs), $L'u$ selection is not applicable to sire of sire selection. A model that considers L'_1y selection and L'_2u selection jointly, for two distinct matrices of full row rank L'_1 and L'_2 , is presented by Famula (4). The mixed model equations derived under this joint selection model are

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z & X'L_1 & 0 \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} & 0 & -L_2 \\ L'_1X & 0 & L'_1VL_1 & L'_1ZGL_2 \\ 0 & -L'_2 & L'_2GZ'L_1 & L'_2GL_2 \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{u} \\ \hat{t}_1 \\ \hat{t}_2 \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \\ L'_1y \\ 0 \end{bmatrix} \quad [6]$$

where $V = \text{var}(y) = (ZGZ' + R)\sigma^2$. If $L'_1X = 0$, which is not unreasonable as sires usually are compared on estimates of transmitting ability corrected for herd, year, and season (fixed effects of the model), it remains to be shown under what circumstances L'_1ZGL_2 is null. If the appropriate model for the data is the groups model of [2] (with Q unknown), then $L'_2 = Q'G^{-1}$, and $L'_1ZGL_2 = L'_1ZQ$. Selection of sires based on comparisons within genetic groups implies that $L'_1ZQ = 0$. Thus, the effect of sire selection (L'_1y selection) can be ignored by using genetic groups (originally intended to account for dam or L'_2u selection) if selection comparisons are within groups.

However, a major argument against consideration of an $L'u$ selection model is that u is an unobservable vector, and as such, selection decisions cannot be based on u directly. Instead, selection is on actual observations (i.e.,

y). The distinction is that some observations may be included in the model while other observations are not available directly. For example, in most sire evaluation models records on female relatives, other than progeny, are not included usually (or attributed to the proper sire through relationships). Thus, a dam's record will not be included in the sire evaluation process unless her own sire is known. And unless a material grandsire model is used, even if her sire were known, a dam record would contribute no information to her son's proof.

A more realistic model would consider joint selection on records in the model for observations and records not in the model. This can be formulated as

$$\begin{bmatrix} L'_1 y \\ L'_2 y_u \end{bmatrix}$$

selection where y_u is a vector of records not included in the model for observations. The $L'_1 y$ selection represents sire of sire selection based on functions of progeny means or previous sire proofs. The $L'_2 y_u$ selection represents dam of sire selection where the dam records used are not included in the evaluation process. Under this form of joint selection, the mixed model equations are

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z & X'L_1 & 0 \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} & 0 & -G^{-1}FL_2 \\ L'_1 X & 0 & L'_1 VL_1 & L'_1 ZFL_2 \\ 0 & -L'_2 F'G^{-1} & L'_2 F'Z'L_1 & L'_2 F'G^{-1}FL_2 \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{u} \\ \hat{t}_1 \\ \hat{t}_2 \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \\ L'_1 y \\ 0 \end{bmatrix} \quad [7]$$

where $F = cov(u, y'_u)$. For most applications F will be a matrix of numerator relationships between the sires in u and the females with records in y_u . Equations [7] as compared to equations [5] show that $L'_2 y_u$ selection (selection on records not in the model) is equivalent to $L'_2 F'G^{-1}u$ selection. Thus, selection on records not included in the model is related to

Henderson's (7) concept of "prior information". Taking this result one step further, equating the genetic groups model of equations [4] (equivalent to $Q'G^{-1}u$ selection) to those of $L'_2 y_u$ selection is equivalent to the genetic groups model with $Q = FL_2$. Thus, as also shown by Quaas and Pollak (13), an appropriate structure for genetic groups may be derived from knowledge of selection decisions.

Substituting Q for FL_2 in equations [7], as in equations [6], shown is that if sire selection decisions are based on comparisons within genetic groups, then $L'_1 ZFL_2 = L'_1 ZQ = 0$. If $L'_1 X = 0$ as well, the standard genetic groups model of [2] is sufficient to account for sire of sire selection and dam of sire selection (based on records not included in the model) when $Q = FL_2$ and sire selection decisions are based on comparisons within groups.

This result can be generalized further. Comparing equivalent linear models shows that if $Qc = FL_2 b$, for any vectors c and b of appropriate order, estimable functions of sire proofs computed under Q will be identical to those computed for FL_2 (if F and L_2 are known). Thus, knowing Q exactly, so that it is identical to FL_2 , is unimportant. Provided vectors c and b exist, estimable functions of sire proofs will be identical under both models. In examples where FL_2 is known and Q is chosen such that c and b do not exist (i.e., $Qc \neq FL_2 b$ for any c and b), consequences of misgrouping sires can be evaluated.

ILLUSTRATION

In this section is a small example to illustrate selection on records not included in the model for observations. The example is taken from Dempfle (1).

Consider two males (animals 1 and 2) and two females (animals 3 and 4) each with n progeny records. On the basis of progeny means (i.e., $\bar{y}_1, \bar{y}_2, \bar{y}_3, \bar{y}_4$) the best male (animal 2) and female (animal 3) were selected and mated to produce two male progeny (animals 5 and 6). Both young sires also have n progeny records (with means \bar{y}_5 and \bar{y}_6). Animals 1 to 4 are assumed to be unrelated individuals from the same population. A simple model for progeny means is

$$\bar{y}_i = \mu + u_i + e_i \quad [8]$$

where y_i ($i = 1, \dots, 6$) is the progeny mean of the i th animal, μ is an unknown constant, u_i is half the breeding value of the i th animal, and $\bar{\epsilon}_i$ is a random residual normally and independently distributed (NID) $[0, \sigma_u^2/n]$. The $\mathbf{u} = \{u_i\}$ ($i = 1, \dots, 6$) and $\text{var}(\mathbf{u}) = \mathbf{G}\sigma_u^2$ where \mathbf{G} is the numerator relationship matrix. For normality and given the specified selection plan

$$E \begin{bmatrix} \bar{y}_1 \\ \bar{y}_2 \\ \bar{y}_3 \\ \bar{y}_4 \\ \bar{y}_5 \\ \bar{y}_6 \end{bmatrix} = \begin{bmatrix} \mu + \delta \\ \mu + \delta \\ \mu + \delta \\ \mu + \delta \\ \mu + a \\ \mu + a \end{bmatrix} \text{ and } E \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \\ u_6 \end{bmatrix} = \begin{bmatrix} -a \\ a \\ a \\ -a \\ a \\ a \end{bmatrix} \quad [9]$$

where $a = \delta(n/(n+k))$ [for $k = \sigma_e^2/\sigma_u^2$] and δ is the expected value of the largest order statistic of a sample of two NID $(\mu, \sigma_u^2 + \sigma_e^2/n)$ random variables (13). Use of order statistics is necessary to determine precisely the expected genetic response under this selection program. For practical problems, with larger numbers of sires and unequal number of observations per sire, this task would be nearly insurmountable, and for this reason the contrived example has been kept simple.

As in the usual practice in sire evaluation, the four males will be evaluated by only their progeny information, i.e., $\bar{y}_1, \bar{y}_2, \bar{y}_5, \bar{y}_6$. Thus, the dam information (\bar{y}_3 and \bar{y}_4) is used for selection but not included in the model for sire evaluation. This can be put in the notation of the previous section as

$$\mathbf{L}'_1 \mathbf{y} = (-1 \quad 1 \quad 0 \quad 0) \begin{bmatrix} \bar{y}_1 \\ \bar{y}_2 \\ \bar{y}_5 \\ \bar{y}_6 \end{bmatrix}$$

$$\mathbf{L}'_2 \mathbf{y}_u = (1 \quad -1) \begin{bmatrix} \bar{y}_3 \\ \bar{y}_4 \end{bmatrix} \text{ and}$$

$$\mathbf{F} = \text{cov}(\mathbf{u}, \mathbf{y}'_u) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ \frac{1}{2}\sigma_u^2 & 0 \\ \frac{1}{2}\sigma_u^2 & 0 \end{bmatrix}$$

To simplify the algebra the example is presented with progeny numbers $n = 40$ and variance ratio $k = 10$. After model [8], equations [7] can be used to compute BLUP of the sire proofs under the joint selection model. As expected $\mathbf{L}'_1 \mathbf{X} = 0$ and $\mathbf{L}'_1 \mathbf{ZFL}_2 = 0$, which simplifies the computations. To simplify computations, relationships among males will be ignored ($\mathbf{G} = \mathbf{I}$). This will not effect estimation of breeding values but will change what effects are contained in the genetic groups (5), as will be shown in the following section. The solution for the sire proofs are

$$\begin{bmatrix} \hat{u}_1 \\ \hat{u}_2 \\ \hat{u}_5 \\ \hat{u}_6 \end{bmatrix} = \begin{bmatrix} .4 & -.4 & 0 & 0 \\ -.4 & .4 & 0 & 0 \\ -.5 & -.5 & .9 & .1 \\ -.5 & -.5 & .1 & .9 \end{bmatrix} \begin{bmatrix} \bar{y}_1 \\ \bar{y}_2 \\ \bar{y}_5 \\ \bar{y}_6 \end{bmatrix} \quad [10]$$

which are unbiased predictors of half the sires' breeding values.

In practice \mathbf{L}'_2 and \mathbf{F} are unknown, and genetic groups are used to account for selection on records not in the model. In this example, an "obvious" way to group sires is by generation such that

$$\mathbf{Q} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}$$

For equations [4] and ignored relationships, solutions for the sire proofs are (under the constraint that $\mu = 0$)

$$\begin{bmatrix} \hat{u}_1 \\ \hat{u}_2 \\ \hat{u}_5 \\ \hat{u}_6 \end{bmatrix} = \begin{bmatrix} .9 & .1 & 0 & 0 \\ .1 & .9 & 0 & 0 \\ 0 & 0 & .9 & .1 \\ 0 & 0 & .1 & .9 \end{bmatrix} \begin{bmatrix} \bar{y}_1 \\ \bar{y}_2 \\ \bar{y}_5 \\ \bar{y}_6 \end{bmatrix} \quad [11]$$

Although actual solutions for the u_i in [11] are different from [10], differences between sire solutions in [11] are identical to those of [10]. Thus, as Quaas and Pollak (13) showed, this definition of genetic groups yields unbiased estimates of the differences between sires. From the previous section, under this definition of genetic groups there exists vectors c and b such that $Qc = FL_2b$. As a result, estimable functions under each model are identical. If, however, Q were chosen to be

$$Q = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$

(which has no intuitive appeal other than not being the same as the previous definition) for which there exists no c and b to satisfy $Qc = FL_2b$, sire proofs would be

$$\begin{bmatrix} \hat{u}_1 \\ \hat{u}_2 \\ \hat{u}_5 \\ \hat{u}_6 \end{bmatrix} = \begin{bmatrix} .9 & 0 & .1 & 0 \\ 0 & .9 & 0 & .1 \\ .1 & 0 & .9 & 0 \\ 0 & .1 & 0 & .9 \end{bmatrix} \begin{bmatrix} \bar{y}_1 \\ \bar{y}_2 \\ \bar{y}_5 \\ \bar{y}_6 \end{bmatrix} \quad [12]$$

Not only are the actual solutions to [12] different from [10], differences between proofs in [12] are biased. For example, $E[\hat{u}_5 - \hat{u}_6] = -.25a$, which should be null. Similarly, if we define Q such that $Q' = (1 \ 1 \ 1 \ 1)$ (i.e., all sires in the same group), differences between sire proofs are biased because c and b do not exist to satisfy $Qc = FL_2b$. Thus, genetic groups cannot be established blindly; instead a grouping structure related linearly to selection practices must be derived to obtain unbiased estimates of proof differences.

DISCUSSION AND CONCLUSIONS

The purpose of this paper is to examine the role of genetic groups in linear models for sire evaluation. Results show the usefulness of genetic groups to account for artificial selection. Without selection the use of genetic groups would be restricted to problems such as the evaluation of Polish Friesians by Stolzman (16), where sires were grouped by country of origin. This paper also presents a correspondence between a model with selection on records not included in the model and a particular form of $L'u$ selection (7). Predicated on the assumptions of the Henderson (7) selection model, this correspondence leads to the conclusion that if sires are selected on comparisons within genetic groups, there will be no bias added to the sire proofs if genetic groups are included. In addition, it is shown that exact representation of the type of dam of sire selection practiced is not important. If a linear relationship between the type of selection on records not in the model and the structure of genetic groups can be found, then the use of genetic groups will yield unbiased estimators of the difference between sire breeding values. Although knowledge of exact selection decisions is not possible in practice, this result can guide definition of genetic groups if an understanding of how dams of sires are selected is available.

One point not addressed is the definition of group effects, i.e., what are the genetic components being estimated by g ? Work in this area has been undertaken by Pollak and Quaas (11); however, as the authors admit, the algebraic expression offered for group parameters is expedient and tentative. Yet, given the modified mixed model equations (line [4]) they derived (which are equivalent to a particular $L'u$ selection model) \hat{g} is equivalent to Henderson's \hat{t} and so are the parameters. Thus, under the correspondence of the groups model (line [1]) with a $Q'G^{-1}s$ selection model the definition of the group parameters can be written as

$$g = (Q'G^{-1}Q)^{-1} E_s[Q'G^{-1}s] \quad [13]$$

where E_s denotes expectation after selection. A derivation of [13] under equivalent models can be found by Famula (5). Therefore, with knowledge of the selection decisions an explicit representation for the group parameters is pos-

sible provided the derivation of $E_s[Q'G^{-1}s]$ is also possible.

Also not addressed in this paper is the definition of the structure of genetic groups, i.e., how to determine Q ? We have shown under what circumstances groups may remain arbitrary. Specific definitions of genetic groups are dependent upon the selection. This paper does suggest goals for determination of Q , the matrix which defines genetic groups. That is, Q should be chosen such that it is related linearly to selection (i.e., $Qc = FL_2b$) or also such that selection of sires is within genetic groups (i.e., $L'_1ZQ = O$). However, to propose a general multipurpose definition for genetic groups is impossible. This does not rule out the possibility of establishing guidelines to aid in definition of genetic groups.

Perhaps the most useful paper in setting guidelines is by Kennedy (9) who demonstrated in a simulated data set with two genetic groups that genetic groups may be ignored without increasing the mean square error of prediction of the estimated sire proofs if the true difference between groups is less than the standard error of the estimated difference. Of course, some bias in the estimator is accepted (8). Thus, if differences between sires is small, ignoring or combining genetic groups is a reasonable alternative provided one realizes the consequences. Conversely, adding genetic groups unnecessarily will not bias prediction of genetic merit, but it will increase the error variance of prediction (4).

An alternative to outlining a strict definition for genetic groups is application of discriminatory analysis or numerical taxonomy. Algorithms for clustering procedures are available in several textbooks (14, 15). One method, with desirable statistical properties is based on the expectation-maximization (EM) algorithm of Dempster et al. (2), which is based on the conditional probability that a given sire belongs to a particular genetic group based on his proof.

The United States Department of Agriculture employs a similar grouping method based on a pedigree index (3). Pedigree indices should reflect effects of selection in previous generations, and, thus, this grouping strategy should aid in controlling the bias caused by selection. Dickinson et al. (3) stated that a "slightly conservative" selection index (using information on

ancestors) is constructed and sires are grouped on the basis of similar index values "encompassing about a fifty-pound range in milk yield". In reference to the selection model presented in this paper, this method should work well at developing a Q , which is related to selection in the population. However, the 22.7-kg range may be too small. For example, if heritability of milk yield is .25, variance of milk records is 900 kg², and accuracy of a pedigree index is equivalent to 10 progeny records, one can compute the probability that a sire's true breeding value is within ± 22.7 kg of his index. As such $\Pr(\text{Index} - 22.7 \leq \text{True Breeding Value} \leq \text{Index} + 22.7) = .05$. This probability seems low if it is to ensure an accurate choice for Q .

Ideally, cows and bulls would be evaluated simultaneously with the numerator relationship matrix of the entire population. Use of the maternal grandsire model proposed by Quaas et al. (12) is a step in this direction. Tying records to sires through relationships reduces the amount of information in y_u , thus eliminating the need for genetic groups. This can be studied further using [13] and various relationship matrices and is illustrated briefly in Famula (5). Unfortunately, however, little attention has been given to applying present computing algorithms to the task of joint cow and sire evaluation.

ACKNOWLEDGMENTS

This study was supported in part by a grant from Eastern Artificial Insemination Cooperative, Inc., Ithaca, NY. The authors would also like to thank Ernst Niebel for his assistance in the necessary translation of reference (1) and an anonymous reviewer for an instructive critique.

REFERENCES

- 1 Dempfle, L. 1978. BLUP unter Verwendung der Verwandtschaftsmatrix. Pages 76-85 in *Zuchtwertschätzung beim Rind*. Habilitationsschrift, München-Weihenstephan.
- 2 Dempster, A. P., N. M. Laird, and D. B. Rubin. 1977. Maximum likelihood from incomplete data via the EM algorithm. *J. Royal Stat. Soc. Ser. B.* 39:1.
- 3 Dickinson, F. N., R. L. Powell, and H. D. Norman. 1976. An introduction to the USDA-DHIA Modified Contemporary Comparison. Pages 1-7 in *The USDA-DHIA Modified Contemporary Comparison*:

- Sire Summary and Cow Index Procedures. Prod. Res. Rep. 165., Agric. Res. Serv., Dep. Agric., Washington, DC.
- 4 Famula, T. R. 1981. Defining genetic groups in dairy sire evaluation. Unpubl. Ph.D. thesis, Cornell Univ., Ithaca, NY.
 - 5 Famula, T. R. 1982. A characterization of group parameters under Henderson's selection model. Mimeo, Dep. Anim. Sci., Univ. California, Davis.
 - 6 Henderson, C. R. 1973. Sire evaluation and genetic trends. Pages 10-41 in Proc. Anim. Breeding Symp. in Honor of Dr. J. L. Lush. Am. Soc. Anim. Sci., Am. Dairy Sci. Assoc., Blacksburg, VA.
 - 7 Henderson, C. R. 1975. Best linear unbiased estimation and prediction under a selection model. *Biometrics* 31:423.
 - 8 Henderson, C. R. 1975. Comparison of alternative sire evaluation methods. *J. Anim. Sci.* 41:760.
 - 9 Kennedy, B. W. 1981. Bias and mean square error in mixed model sire evaluation when genetic group effects are ignored or included. *J. Dairy Sci.* 64: 689.
 - 10 Pearson, K. 1903. Mathematical contributions to the theory of evolution. XI. On the influence of natural selection on the variability and correlation of organs. *Philos. Trans. Royal Soc. London A* 200:1.
 - 11 Pollak, E. J., and R. L. Quaas. 1983. Definition of group effects in sire evaluation models. *J. Dairy Sci.* 66 (In press).
 - 12 Quaas, R. L., R. W. Everett, and A. C. McClintock. 1979. Maternal grandsire model for dairy sire evaluation. *J. Dairy Sci.* 62:1648.
 - 13 Quaas, R. L., and E. J. Pollak. 1981. Modified BLUP equations for sire models with groups. *J. Dairy Sci.* 64:1868.
 - 14 Rao, C. R. 1973. Linear statistical inference and its application. John Wiley and Sons, Inc., New York, NY.
 - 15 Sokal, R. R., and P.H.A. Sneath. 1963. Principles of numerical taxonomy. W. H. Freeman and Co., London, Great Britain.
 - 16 Stolzman, E. M. 1979. Testing different strains of Friesian cattle in Poland. Fifth Annu. Rep., Inst. Zootech., Cracow, Poland.
 - 17 Thompson, R. 1979. Sire evaluation. *Biometrics* 35:339.