

## C. R. Henderson: Contributions to Predicting Genetic Merit

L. R. SCHAEFFER

Centre for Genetic Improvement of Livestock  
University of Guelph  
Guelph, ON, Canada N1G 2W1

### ABSTRACT

The contributions of C. R. Henderson to the genetic evaluation of livestock have been widely accepted, utilized, and enhanced by animal breeders and statisticians. Not well known are the possible alternatives to BLUP that have been suggested by Henderson, such as biased predictors and Bayesian methodology to incorporate prior information. A search for rapid methods of inverting dominance and additive by additive genetic relationship matrices has taken place since Henderson first described his method for computing the inverse of the additive genetic relationship matrix. Accounting for inbreeding and selected base populations continues to be a problem. There is a need to derive accurate descriptions of selection processes and the appropriate selection model in order to solve the problems of cow culling, nonrandom association of sires with herd-year-seasons, and preferential treatment. Henderson solved many animal breeding problems and left hints for solving others, but there are still many difficult problems to be tackled that will have to be resolved without the availability of his generous advice.

(Key words: genetic merit, Henderson, symposium)

Abbreviation key: HYS = herd-year-season, MME = mixed model equations, PREF = preferential treatment.

### INTRODUCTION

C. R. Henderson was significantly inspired by the 1933 paper of J. L. Lush, "The Bull

Index Problem in the Light of Modern Genetics" (36). This paper provided the challenges to Henderson for the derivation of BLUP, his three methods of variance component estimation, and the selection bias methods. The Lush paper was characterized by detailed assumptions associated with particular models or methods. This trait was inherited by Henderson as all of his publications and writings contain precise and consistently stated assumptions and conditions under which the models and methodology are applicable.

The objective of this paper is to outline the major achievements of Henderson in genetic evaluations of animals with particular emphasis on problems that remain to be resolved. Enhancements to these methods provided by other researchers are noted.

### BLUP and Mixed Model Equations

*Historical Development.* The discovery of BLUP and mixed model equations (MME) by Henderson was the result of a problem he was assigned in a statistics class (39). Henderson solved the problem using Lush's most probable producing ability methods. This led to his thesis work, which attempted to combine the theories of least squares and selection index in the estimation of general, specific, and maternal abilities in swine (9). The MME were first presented in 1949 (10, 11). However, the proof that solutions for fixed effects were the same as those obtained from generalized least squares equations did not occur until 1956 with Searle's help and was not published until 1959 (33). Proof that solutions for random effects were the same as the selection index criterion with generalized least squares solutions for fixed effects used as means was also derived in 1956 but again not published until 1963 almost 2 yr after it was originally submitted (13). Goldberger (8) published the term BLUP in 1962. This history is summarized in

Received June 30, 1990.

Accepted March 20, 1991.

the 1973 Lush Symposium paper by Henderson (15).

*The Prediction Problem.* Let a general mixed linear model be

$$y = X\beta + Zu + e \quad [1]$$

where  $y$  is the observation vector,  $X$  is a known, fixed,  $n \times p$  matrix with rank  $r$ ,  $\beta$  is an unknown, fixed vector,  $Z$  is a known, fixed matrix,  $u$  and  $e$  are nonobservable random vector variables with null means, and the following variance-covariance structure:

$$V \begin{bmatrix} u \\ e \end{bmatrix} = \begin{bmatrix} G & 0 \\ 0 & R \end{bmatrix} \quad [2]$$

where  $G$  and  $R$  are known and nonsingular. Animal breeders nearly always assume that a linear model is appropriate, and in many situations this is true. However, the possibility that nonlinear models may be better in some situations must not be ignored (2). Nonlinear methodology may be complex to implement, and therefore, the definition of "better" and a measure of how much "better" will determine how readily, if at all, nonlinear methods are implemented.

To obtain BLUP, we would like to predict a linear function of  $\beta$  and  $u$ , say  $K'\beta + M'u$ , by a linear function of  $y$ ,  $F'y$ , such that the expectation of  $F'y$  equals the expectation of  $K'\beta + M'u$  and that the variances of prediction error in  $(F'y - K'\beta - M'u)$  are minimized.  $K'\beta$  is assumed to be estimable. If the derivatives of a function of the variance of prediction error matrix, incorporating a LaGrange multiplier matrix to force unbiasedness, are equated to zero, then a solution for  $F$  is obtained. The predictor is

$$F'y = K'\hat{\beta} + M'\hat{u}$$

where  $\hat{\beta}$  and  $\hat{u}$  are solutions to Henderson's MME:

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

The general matrix representation in [3] encompasses sire models, animal models, and multiple-trait models, each of which have been presented and discussed by Henderson (15, 16, 17, 21, 23, 29, 34). Although the term "animal model" became popular from Quaas and Pollock (45), Henderson certainly wrote about and discussed animal models in his animal breeding courses at Cornell in the early 1970s and earlier (12). A simple sire model, with sires assumed unrelated, was the first practical application of [3] in the Northeast AI Sire Comparison. Approximately 1200 Holstein AI sires were evaluated by a series of computer programs written by Henderson for a 128 kb IBM 360 computer.

Equations [3] are formed without the need to know the distribution of  $y$ . For hypothesis testing or estimation of variance components by maximum likelihood or REML, normality is commonly assumed. However, the (co)variance matrices,  $G$  and  $R$ , are assumed known, nonsingular, and positive definite. Gianola et al. (6) have described methods to use when  $G$  and  $R$  are not known, and Henderson has given computational algorithms for the case when  $G$  or  $R$  or both are singular (15, 23).

Computing aspects of constructing MME always intrigued Henderson. He introduced the concept of equivalent models in an attempt to simplify computing or understanding of various models. Two models are said to be equivalent if they give the same expectation and covariance matrix of  $y$  (25). An example is the reduced animal model which is useful for situations where there are many offspring per mating, such as in swine, or where very few individuals are kept as parents of future animals (45). Henderson also described the use of a model having a sire by herd interaction effect to account for an environmental correlation among half-sibs in the same herd (15). Schaeffer and Henderson (47) derived BLUP for models where  $\text{Cov}(u, e)$  is not null, and showed that for such models an equivalent model could be written in which the covariance matrix between the random vector of the model and the residual vector is null.

*Incorporating Prior Information.* Another prediction problem considered by Henderson is to predict  $K'\beta + M'u$  by  $F'y + S'\beta^*$ , such that

$$\begin{aligned} E(K'\beta + M'u) &= E(F'y + S'\beta^*) \\ &\text{and diagonals of} \\ V(K'\beta + M'u - F'y - S'\beta^*) \\ &\text{are minimized.} \end{aligned}$$

The vector  $S'\beta^*$  is a prior unbiased estimate of  $S'\beta$  assumed to be estimable and independent of  $y$  and  $V(\beta^*) = P$ , which is nonsingular. The resulting MME from this problem are

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

Thus, previous estimates of fixed effects, such as breed differences, differences in ages at calving, and sex differences could be incorporated. Note that if  $P^{-1}$  is null, then the solution for  $\beta$  is derived totally from the data, and if the elements of  $P^{-1}$  are large, then the solutions become similar to the prior values.

Gianola and Fernando (4) derived similar equations in a Bayesian context in which there is no need to distinguish between fixed and random factors. Their equations were

$$\begin{bmatrix} X_1'R^{-1}X_1 & X_1'R^{-1}X_2P & X_1'R^{-1}Z \\ PX_2'R^{-1}X_1 & PX_2'R^{-1}X_2P+P & PX_2'R^{-1}Z \\ Z'R^{-1}X_1 & Z'R^{-1}X_2P & Z'R^{-1}Z+G^{-1} \end{bmatrix} \begin{bmatrix} \beta_1^* \\ \alpha_2^* \\ u^* \end{bmatrix} = \begin{bmatrix} X_1'R^{-1}y \\ PX_2'R^{-1}y \\ Z'R^{-1}y \end{bmatrix} \quad [6]$$

where  $X$  has been partitioned into  $(X_1 X_2)$ ,  $\beta' = (\beta_1' \beta_2')$ , and  $P = \beta_2\beta_2'$ , so that  $\beta_2^* = P\alpha_2^*$ , and  $\beta_2$  is a prior value of  $\beta_2$ . Henderson noted that the solutions for  $\beta_2^*$  resulting from [6] were proportional to  $\beta_2$ , the vector of priors used to form  $P$ , and this property was deemed undesirable.

Bayesian methods (3) may be preferable to the use of Equations [6] and may yield improved solutions over BLUP in the sense of smaller mean squared error. Biased estimators may also be formed by simply adding a constant to the diagonal elements of the MME for fixed factors in the model. If biased estimators

$$\begin{bmatrix} X'R^{-1}X + P^{-1} & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y + P^{-1}\beta^* \\ Z'R^{-1}y + G^{-1}u^* \end{bmatrix} \quad [5]$$

The right sides of Equations [5] also contain prior information about  $u$ . If no prior information exists for  $\beta^*$  and  $u^* = 0$ , then [5] reduces to [3], the usual MME.

Equations [5] could be useful for within-herd evaluations in which breeding values of ancestors from an analysis of data from all herds could be incorporated into the equations for a given herd. This differs from the approach of Henderson who used selection index theory (19) to derive modifications to both the coefficient matrix and the right sides of [3] for use in within-herd evaluation of dairy cows (51).

*Biased Prediction.* In 1987 Henderson emphasized the possibility that unbiased predictors may not be essential (27, 28); that is, a biased predictor may exist that has smaller mean squared error than an unbiased predictor. Henderson outlined BLUP (23) and obtained the following equations:

are to be used, then methods for measuring bias and computing mean squared errors are necessary (30). If the bias in an estimator can be shown to be small, then publication of biased predictors of genetic merit of individuals should be acceptable to producers but may have legal ramifications.

#### Numerator Relationship Matrices

*Historical Development.* The second most important discovery of Henderson was a rapid method for computing the inverse of the numerator relationship matrix from a list of individual pedigrees from a noninbred popula-

tion (20). The Northeast AI Sire Comparison model was changed in February 1975 to incorporate relationships among bulls based on their sire and maternal grandsire (MGS) ancestors. Although the use of additive genetic relationships was a great step forward in genetic evaluations, several new problems arose from it (35, 41, 43).

The first problem encountered was the effect of relationships on genetic group solutions. Group differences for AI studs and years of entry were smaller in magnitude in the model with relationships than in the model without. Henderson contended that if there had been complete pedigree information on each bull so that each bull could be traced to the same "fixed" base generation, then group differences should have been close to zero. However, because the relationships were based only on sire and MGS information, Pollak and Quaas (41) showed that the group solutions reflected differences in selection differentials on the female side of the pedigrees.

Accounting for inbreeding in the inverse of the relationship matrix was solved by Quaas (42) with an algorithm that required the pedigrees to be ordered chronologically. Tier (50) devised a rapid method of computing the numerator relationships (before inversion) that includes the inbreeding coefficients. The algorithm will be very helpful in monitoring inbreeding rates, especially in small populations. Once the inbreeding coefficient is known, then computing the correct inverse elements follows directly.

The rate of convergence of iterative schemes to obtain solutions to the mixed model equations changed with the incorporation of relationship matrices. Without the relationship matrix, solutions for sires were commonly achieved after 15 rounds of iteration from initial starting values of zero. Due to the additional off-diagonal elements in the sire by sire part of the MME from the inverse of the relationship matrix, the number of rounds of iterations needed increased. The number of rounds of iterations to reach convergence with the animal model also increased because the diagonal elements of the MME were no longer dominant to the off-diagonals. Computing strategies for solving MME have changed since Henderson's original sire model programs. Tridiagonalization, sparse matrix tech-

niques, iteration on data, and relaxation factors are procedures that animal breeders must utilize today (38). In addition, there may be clever manipulations of the MME, such as by Quaas (44), to simplify computations.

*Genetic Grouping of Animals.* The biggest problems in utilizing additive genetic relationships, in practice, are the inability to trace the pedigree of every animal back to a common base generation, and to ensure that the base population animals represent an unselected group of individuals. Henderson showed that if the base population animals resulted from selection, then the inverse of the relationship matrix should be modified to account for selection (24, 31). However, Henderson found that solutions from MME ignoring selected base animals were the same as solutions that accounted for selected base animals, but he urged extreme caution in interpreting the solutions. McAllister et al. (37) have utilized the modified inverse of the relationship matrix in a study of a designed experiment.

The lack of complete pedigrees on every animal back to the same common, unselected base generation necessitates assigning animals to genetic groups. Westell et al. (52) and Quaas (43) proposed the use of phantom parents which can then be grouped by year of birth of their progeny or some other arbitrary manner. Incorporating the phantom parent groups into the MME is accomplished with a slight modification to Henderson's rules for computing the inverse of the relationship matrix. For each animal we need  $m$ , the number of unknown parents, then  $x = 4/(m + 2)$ , and we add  $x$  to the animal's diagonal element,  $-x/2$  to the animal by parent (or parent group) elements, and  $x/4$  to the parent diagonal and off-diagonal elements. The solutions for the animal additive genetic effects in the MME are estimates of functions of the group effects for the animals plus their additive genetic effects (as deviations from the group mean). The function of group effects included in each animal's solution depends on that animal's particular pedigree.

If missing parents are grouped according to sex and year of birth of their offspring, and if unknown dams and sires are grouped separately, then often some of the group equations will be the same or nearly the same, resulting in a singular coefficient matrix in the MME. Deter-

mining which functions of the group effects are estimable may not be simple. Kennedy (personal communication) has proposed staggering the year groups for unknown sires and dams to break up the confounding of phantom group equations. Assuming different generation intervals for sire and dam may also eliminate this problem (Van Vleck, personal communication). For animals with both parents unknown, both sire and dam could also be allocated to the same phantom group (Wiggans, personal communication).

A new method for dealing with selected base animals and incomplete pedigrees follows. According to Henderson (31), the modified inverse of the relationship matrix and the phantom grouping concept of Westell et al. (52) were shown to be directly related. The new proposal is to let each phantom animal be its own group, and assumed to have only one progeny each. In addition, every animal is traced back to a base generation in which all animals are assumed to be unselected (with

phantom animals inserted wherever necessary). To account for selected phantom animals after generation 0, the rules of Quaas (43) in forming the Henderson modified inverse of the relationship matrix need to be used. Each phantom animal would be treated as a real animal with both of its parents being phantom animals. Thus, phantom animals, in this proposal, would be random effects.

If the base generation is 50 generations removed from the current animals, then the pedigree file could contain a very large number of phantom animals. Fortunately, there is no need to construct the inverse for all real and phantom animals in the pedigree file. Phantom animals can be absorbed into real animals. This leads to a simple set of rules for creating the appropriate "absorbed", modified inverse of the relationship matrix for real animals only. For each real animal, however, with one or both parents unknown, the number of generations removed from generation 0 must be known. Consider the example pedigree in Figure 1; phantom animals are denoted by numbers and real animals are denoted by letters. Animals 5 through 16 form generation 0 in this example. Now construct the following table of information on each real animal where  $m$  is the number of unknown parents and  $t$  is the number of generations removed from the base.

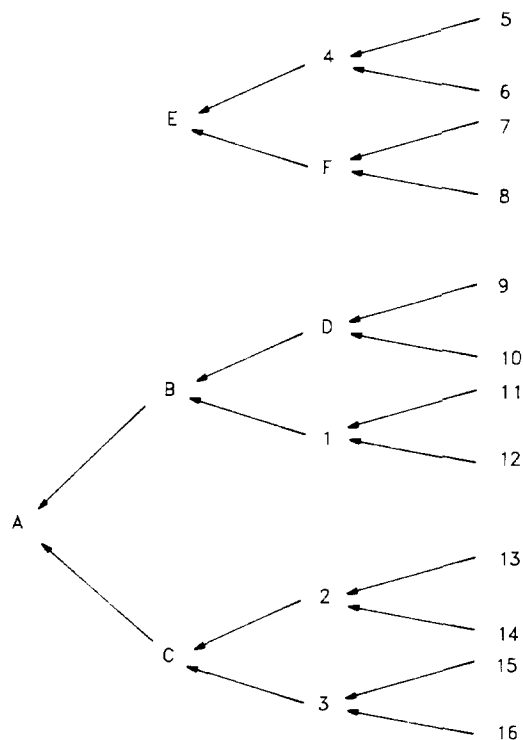


Figure 1. Example pedigree of animals traced to the same base population.

Animal	Sire	Dam	m	t
A	B	C	0	3
B	D	...	1	2
C	...	...	2	2
D	...	...	2	1
E	...	F	1	2
F	...	...	2	1

The rules for calculating the "absorbed", modified inverse are as follows.

1. When  $m = 0$ , then  $x = 2$ , and  $x$ ,  $-x/2$ , and  $x/4$  are added to the appropriate positions of the inverse as given before.
2. When  $m = 2$ , then  $x = 1$  and compute  $w = (2^t - 1)/(2^{t+1} - 1)$  and add  $(x - w)$  to the diagonal of that animal.

3. When  $m = 1$ ,  $x = 4/3$ , and compute  
 $p = [x - (2^{t-1} - 1)/(2^t - 1)]^{-1}$ ,  
 then  
 add  $(x - 4p/9)$  to the  
 animal's diagonal,  
 add  $(-x/2 + 2p/9)$  to the ani-  
 mal by known parent  
 off-diagonal,  
 and add  $(x/4 - p/9)$  to the known  
 parent's diagonal.

The resulting inverse for the real animals in the example is given.

	A	B	C	D	E	F
A	2	-1	-1	0	0	0
B	-1	25/18	1/2	-4/9	0	0
C	-1	1/2	15/14	0	0	0
D	0	-4/9	0	8/9	0	0
E	0	0	0	0	8/9	-4/9
F	0	0	0	0	-4/9	8/9

The difficult part of this method is defining the base generation and determining the number of generations each animal is removed from the base. Assumptions about average generation intervals for sires of bulls, sires of cows, dams of bulls, and dams of cows will need to be made as in Westell et al. (52). For North American dairy cattle populations, an appropriate base generation may be animals born before 1945. Note that as  $t$ , the number of generations removed from the base, increases  $w = (2^t - 1)/(2^{t+1} - 1)$  approaches the value .5. Thus, any new animals that appear in the data in 1990 might simply be assumed to be many generations removed from the base rather than actually trying to count the number of generations. Further study of this method is needed as there might be problems in iterating a solution because some of the diagonal elements of the relationship matrix inverse may end up being less than one as in the previous example.

**Nonadditive Genetic Effects.** The development of animal models for the estimation of nonadditive genetic effects was given by Henderson (26, 32). The model is

$$y = X\beta + Zg + e \quad [7]$$

where  $y$ ,  $X\beta$ ,  $Z$ , and  $e$  are as before for [1], and  $g$  is the vector of total genetic effects of

animals, which can be defined, for example, as the sum of effects of additive ( $a$ ), additive by additive ( $aa$ ), dominance ( $d$ ), dominance by dominance ( $dd$ ), and additive by dominance ( $ad$ ) gene actions, each assumed to be independent of the others. The covariance matrix of  $g$  in a noninbred population in linkage equilibrium is  $M$ , where

$$M = V(a) + V(aa) + V(d) + V(dd) + V(ad)$$

and if  $A$  is the numerator relationship matrix and  $D$  is the dominance genetic relationship matrix, then

$$M = A\sigma_a^2 + A \# A\sigma_{aa}^2 + D\sigma_d^2 + D \# D\sigma_{dd}^2 + A \# D\sigma_{ad}^2 \quad [8]$$

where  $\#$  is the Hadamard product, and the MME are

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + M^{-1} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{g} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix} \quad [9]$$

Having calculated a solution for  $\hat{g}$ , then

$$\begin{aligned} \hat{a} &= AM^{-1}\hat{g}, \\ \hat{d} &= DM^{-1}\hat{g}, \\ aa &= (A \# A)M^{-1}\hat{g}, \\ dd &= (D \# D)M^{-1}\hat{g}, \text{ and} \\ ad &= (A \# D)M^{-1}\hat{g}. \end{aligned}$$

The problem with this model is the calculation of  $M^{-1}$ , because the simple rules of Henderson for computing  $A^{-1}$  do not apply to  $M^{-1}$ . Thus, a strong arm inversion of  $M$  would be needed or equivalent MME that do not require the inverse of  $M$  must be constructed (23).

Attempts have been made to estimate dominance genetic effects and the dominance genetic variance from dairy cattle data in which large full-sib families have been generated by embryo transfer (49). The inverse of  $D$  could be calculated by making  $D$  a block diagonal matrix where the blocks are small enough to

be inverted directly. As of yet there are no simple rules for creating the inverse of  $D$  from a list of pedigrees. However, Smith and Maki-Tanila (48) have developed a procedure derived from a gametic relationship matrix which allows for inbreeding. Similarly, Chang et al. (1) have developed procedures for computing the inverse of  $(A \# A)$  directly, for an  $A$  based on sire and MGS relationships. There will likely be efficient algorithms developed for inverting covariance matrices for various types of gene action before we have suitable data on which to apply the techniques. However, the assumption of linkage equilibrium needs to be relaxed and appropriate procedures developed to compute the correct covariance matrices. All of these developments have been motivated by Henderson's discovery of a way to invert the additive genetic relationship matrix.

#### Selection Bias

An important area of future research for animal breeders is the development of evaluation procedures that adequately account for different kinds of selection bias. Henderson first published on these problems in 1965 (14), later in 1973 (15), and more completely in 1975 and 1982 (18, 22). Selection is defined as any process, natural or artificial, which prevents certain individuals from surviving or reproducing, or in allowing others to do so. A key to understanding bias caused by selection is to describe accurately the selection process.

According to Henderson (18), selection can be on elements of  $y$ ,  $u$ ,  $e$ , or combinations of these. Selection on elements of  $u$  and  $e$  are difficult to conceptualize because  $u$  and  $e$  are nonobservable vectors. Selection can also be direct or indirect through a correlated variable, say  $w$ , or a combination of these. Selection can be described by linear functions of  $y$ , or  $w$ , or  $y$  and  $w$  or as nonlinear functions. Describing the selection process, therefore, requires defining the variable or variables upon which decisions are made and the rules or procedures for making the decisions. This is not an easy task.

Using the notation of Henderson (18), define  $w$  as the vector of information on all candidates available for selection such that  $w$  is

correlated to  $y$ ,  $u$ , or  $e$ , as follows:

$$E \begin{bmatrix} y \\ u \\ e \\ w \end{bmatrix} = \begin{bmatrix} X\beta \\ 0 \\ 0 \\ d \end{bmatrix}$$

and

$$V \begin{bmatrix} y \\ u \\ e \\ w \end{bmatrix} = \begin{bmatrix} V & ZG & R & C_{yw} \\ & G & 0 & C_{uw} \\ & & R & C_{ew} \\ \text{symmetric} & & & C_w \end{bmatrix}$$

The matrices  $G$ ,  $R$ ,  $C_{yw}$ ,  $C_{uw}$ ,  $C_{ew}$ , and  $C_w$  contain known parameter values prior to selection.

Henderson chose to describe selection processes as linear functions of the culling variate,  $w$ , say  $L'w$ . The examples covered in the papers of Henderson provide various  $L'$  matrices for specific problems, but Henderson never gave explicit guidelines for the formation of  $L'$  in a general context. Most of the examples of  $L'$  were such that the rows of  $L'$  would add to zero, and therefore, each row would describe a "differential" in the elements of  $w$  between a selected individual or group of individuals from those individuals that were culled. To illustrate, suppose that  $w$  is a vector of length seven and that the first three individuals are selected over the other four. An  $L'$  that would reflect these decisions is

$$L' = \frac{1}{4} \begin{bmatrix} 4 & 0 & 0 & -1 & -1 & -1 & -1 \\ 0 & 4 & 0 & -1 & -1 & -1 & -1 \\ 0 & 0 & 4 & -1 & -1 & -1 & -1 \end{bmatrix} \quad [10]$$

If the three selected animals were selected as a group, rather than individually, then  $L'$  might be

$$L' = (1/12) [4 \ 4 \ 4 \ -3 \ -3 \ -3 \ -3]. \quad [11]$$

If  $w$  represented individual cows and an analysis was made via a sire model, then use of either [10] or [11] to account for selection might lead to very similar estimated breeding

values for sires. If, however, an analysis was by an animal model, then use of [10] versus [11] could lead to different estimates of breeding values for cows because [10] would account for individual differentials while [11] would adjust for an average differential for the group of three selected cows. Thus, the selection rule and model of analysis could be critical to the removal of selection bias from estimated breeding values.

Henderson adopted the selection model of Pearson (40) and assumed multivariate normality to derive the results presented in 1985 (18). To continue with this derivation, let

$$\begin{aligned} B &= C_{yw}L, \\ B_u &= C_{uw}L, \\ B_e &= C_{ew}L, \text{ and} \\ H &= L'C_wL. \end{aligned}$$

Henderson wrote "conditional" expectations in the sense of conditional given selection on  $w$ , which is different from conditional given  $L'w$ . The term "conditional" should not be used in this context. The expectation and covariance matrices accounting for selection on  $w$  (denoted by  $E_s$  and  $V_s$ ) are

$$E_s \begin{bmatrix} y \\ u \\ e \\ L'w \end{bmatrix} = \begin{bmatrix} X\beta \\ 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} BH^{-1}\alpha \\ B_uH^{-1}\alpha \\ B_eH^{-1}\alpha \\ \alpha \end{bmatrix} \quad [12]$$

$$\text{and} \quad V_s \begin{bmatrix} u \\ e \\ L'w \end{bmatrix} = \begin{bmatrix} G - B_uH_0B'_u & -B_uH_0B'_e & B_uH^{-1}H_s \\ \text{symmetric} & R - B_eH_0B'_e & B_eH^{-1}H_s \\ & & H_s \end{bmatrix} \quad [13]$$

where  $H_0 = H^{-1}(H - H_s)H^{-1}$ ,  $\alpha$  and  $H_s$  are the mean and variance of the selected sample. Generally,  $\alpha$  and  $H_s$  will be difficult to obtain without assuming a certain type of selection (such as truncation selection) and knowing the intensities of selection.

The mixed model equations derived by Henderson (18) using [12] and [13] were

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z & X'R^{-1}B_e \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} & Z'R^{-1}B_e - G^{-1}B_u \\ B'_eR^{-1}X & B'_eR^{-1}Z - B'_uG^{-1} & B'_eR^{-1}B_e + B'_uG^{-1}B_u \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{u} \\ \hat{t} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \\ B'_eR^{-1}y \end{bmatrix} \quad [14]$$

The remarkable feature of [14] is that  $H_s$  and  $\alpha$  do not appear, and thus, are not needed to obtain unbiased predictions. However, the solutions,  $\hat{t}$ , provide an estimate of  $H^{-1}\alpha$ .

A problem with the Pearson model (40) of selection is the nature of the assumed sampling process. Suppose there are two cows (A and B) with records and they are ranked on the basis of their records as  $A > B$ . If this situation were to be simulated under the Pearson model, then only those samples in which  $A > B$  were observed would be accepted in a repeated sampling study. This is in contrast to a sampling scheme where sometimes the ranking is  $A > B$  and other times  $B > A$ . Thus, the ranking of animals is random, and therefore the associ-

ated design matrices in the model would be randomly determined (5, 7). The latter sampling process seems more realistic than that assumed by the Pearson model and as used by Henderson, but given that there is usually only

TABLE 1. Example data for cow culling problem.

Cow	bST Treatment	Milk yields (kg)	
		First lactation	Second lactation
1	Yes	5800	...
2	Yes	6100	6200
3	Yes	5700	...
4	No	6300	6200
5	No	5900	...



one sample of data, then the sampling process as visualized under repeated sampling could be debated. The following material is based on Henderson (18) and the Pearson model of selection. For alternatives to the Pearson model see work by Gianola et al. (5, 7).

*Cow Culling.* Consider the case of five contemporary cows having first lactations of which two were allowed to have a second lactation. Three of the cows were similarly treated with bST and two were not. The data are given in Table 1. The model could be (ignoring other environmental effects)

$$y_{ijk} = \mu + b_i + a_j + p_j + e_{ijk}$$

where  
 $y_{ijk}$  is milk yield,  
 $\mu$  is a contemporary group mean (fixed),  
 $b_i$  is an effect due to bST treatment (fixed),  
 $a_j$  is an animal additive genetic effect (random),  
 $p_j$  is a cow's permanent environmental effect (random), and  
 $e_{ijk}$  is a residual effect (random).

Let  $a$ ,  $p$ , and  $e$  represent the vectors of additive genetic, permanent environmental, and residual effects, respectively, then

$$V(a) = A\sigma_a^2$$

where  $A = (1/16)$  
$$\begin{bmatrix} 16 & 8 & 2 & 3 & 8 \\ 8 & 16 & 2 & 3 & 8 \\ 2 & 2 & 17 & 4 & 12 \\ 3 & 3 & 4 & 16 & 2 \\ 8 & 8 & 12 & 2 & 18 \end{bmatrix}$$

and  $\sigma_a^2$  is the additive genetic variance,

$$V(p) = I\sigma_p^2, \text{ and } V(e) = I\sigma_e^2$$

with  $\sigma_p^2$  being the permanent environmental variance and  $\sigma_e^2$  being the residual variance. If heritability is .25 and repeatability is .5, then

$$k_a = \sigma_e^2/\sigma_a^2 = 2 \text{ and } k_p = \sigma_e^2/\sigma_p^2 = 2.$$

Selection is assumed to have been on first lactations, so that  $w = y$ , and selection is of the  $L'y$  type. An appropriate  $L'$  similar to [10] may be

$$L' = (1/4) \begin{bmatrix} -1 & 4 & -1 & -1 & -1 & 0 & 0 \\ -1 & -1 & -1 & 4 & -1 & 0 & 0 \end{bmatrix}$$

for  $y' = (5800 \ 6100 \ 5700 \ 6300 \ 5900 \ 6200 \ 6200)$ , so that

$$L'y = \begin{bmatrix} 175.0 \\ 425.0 \end{bmatrix}.$$

To derive the appropriate MME for  $L'y$  type of selection, we have

$$\begin{aligned} B_e &= \text{Cov}(e, L'y) = RL, \\ B_u &= \text{Cov}(u, L'y) = GZ'L. \end{aligned}$$

Then

$$\begin{aligned} X'R^{-1}B_e &= X'L, \quad Z'R^{-1}B_e - G^{-1}B_u = 0, \\ B_e'R^{-1}B_e + B_u'G^{-1}B_u &= L'(ZGZ' + R)L = L'VL, \text{ and} \\ B_e'R^{-1}y &= L'y. \text{ The equations to solve are} \end{aligned}$$

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z & X'R^{-1}Z & X'L \\ Z'R^{-1}X & Z'R^{-1}Z + A^{-1}\lambda_a & Z'R^{-1}Z & 0 \\ Z'R^{-1}X & Z'R^{-1}Z & Z'R^{-1}Z + I\lambda_p & 0 \\ L'X & 0 & 0 & L'VL \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{a} \\ \hat{p} \\ \hat{t} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \\ Z'R^{-1}y \\ L'y \end{bmatrix}$$

where  $\lambda_a = \sigma_a^{-2}$  and  $\lambda_p = \sigma_p^{-2}$ .

The solutions for this example are given in Table 2, along with solutions ignoring selection and with solutions using an  $L'$  as in [11] for groups of selected animals. The solutions accounting for selection do not result in any changes in ranking and only minor differences in absolute value from the solutions ignoring selection. The solutions for  $t$  which estimate  $H^{-1}\alpha$ , are modifications of  $L'y$  adjusted for differences in the bST versus non-bST groups and for relationships among cows. The solutions with the two methods of accounting for selection are less different from each other than from the solutions that ignore selection.

Simulation with the appropriate sampling process is needed to determine which method of accounting for cow culling should be used.

In this example,  $L'X$  was not null because one animal was selected from the bST-treated group and one was from the untreated group.  $L'X$  would also not be null if covariates had been included in the model. Cow culling could be ignored if  $L'X$  was null, for example, if selection were entirely within fixed effects subclasses.

*Association Between Herds and Sires.* Many sire evaluation models have included herd-year-season (HYS) effects, which can logically be considered a random variable. Because of the possibility that better sires are associated with better herds, the HYS effects

TABLE 2. Solutions (in kilograms) to cow culling example by two methods of accounting for selection (individual basis and group basis) and by ignoring selection.

Solution for	Ignoring selection	Accounting for selection	
		Individual basis	Group basis
bST or non-bST	-226.92	-142.94	-159.21
Additive genetic			
Cow 1	-12.27	-18.09	-16.96
Cow 2	30.42	30.43	30.43
Cow 3	-62.04	-64.86	-64.32
Cow 4	41.34	51.54	49.56
Cow 5	-59.73	-60.34	-60.22
Permanent environment			
Cow 1	-23.07	-35.96	-33.46
Cow 2	62.32	61.07	61.31
Cow 3	-39.82	-53.70	-51.01
Cow 4	50.13	64.68	61.94
Cow 5	-49.56	-36.19	-38.73
$t_1$	...	298.96	468.15
$t_2$	...	328.34	...

TABLE 3. Example data for sire by herd-year-season (HYS) association problem.

HYS	Sire number	Cow number	First lactation yield (kg)
1	1	5	6100
1	1	6	6300
1	2	7	5800
2	1	8	6500
2	2	9	6000
2	4	10	5000
3	1	11	6400
3	1	12	6200
3	4	13	5200
4	2	14	5900
4	2	15	5700
4	3	16	5600
5	3	17	5500
5	3	18	5800
5	4	19	5400
5	4	20	5100

have been treated as fixed effects in the MME. Henderson has shown that the expectations of sire solutions were equal to the expectations of true sire values under the selection model. For countries with small herd sizes, there are many HYS subclasses with only three or less first lactation heifers, and with HYS effects as fixed, the majority of the HYS subclasses would not contribute any information to the sire equations after absorption of HYS equations. In such countries one might be able to assume that an association between sires and HYS effects does not exist and simply let HYS effects be random. If the association does exist, then HYS effects could still be random, but the user would have to accept that sire solutions would be biased and perhaps have smaller mean squared error. This example illustrates that models developed for one country or situation may not be directly transportable to other situations.

Henderson defined the association between sires and HYS effects as an association between true sire breeding values and true HYS effects, both of which are nonobservable. In terms of the linear model, this would be  $L'u$

type of selection, where  $u$  in this case would be the vector of true HYS effects. Let the model equation be

$$y = Xb + Za + Fh + e$$

where

$y$  is the vector of observations,  
 $b$  is a vector of fixed effects,  
 $a$  is a random vector of animal (sire) additive genetic effects,  
 $h$  is a random vector of HYS effects,  
 $e$  is a random vector of residual effects, and

$X$ ,  $Z$ , and  $F$  are corresponding incidence matrices.

In the no selection situation,

$$V(a) = A\sigma_a^2,$$

$$V(h) = I\sigma_h^2, \text{ and}$$

$$V(e) = I\sigma_e^2.$$

If  $w = h$ , then  $L'$  could be formed to consist of  $s$  rows for the  $s$  sires with progeny in those HYS subclasses, and each row would reflect the 'differential' in true HYS effects of those HYS subclasses in which each sire had progeny from those subclasses in which a sire had no progeny. Each sire should have at least one HYS subclass in which it has no progeny for this definition of  $L'$  to be formed. There will likely be sires that have progeny in exactly the same HYS subclasses, and their two rows of  $L'$  would be identical. Likewise, two sires might have progeny in exactly the opposite HYS subclasses, so that their two rows of  $L'$  are the same except for sign. Only one of the two rows needs to be included in  $L'$ . If all sires had progeny in all HYS subclasses, then there would be no need to account for a non-random association of sires with HYS.

For  $L'u$  type of selection and the model above, Henderson recommended using the following equations:

$$\begin{bmatrix} X'X & X'Z & X'F & 0 \\ Z'X & Z'Z+A^{-1}k_s & Z'F & 0 \\ F'X & F'Z & F'F+Ik_h & -L \\ 0 & 0 & -L' & L'Lk_h^{-1} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{s} \\ \hat{h} \\ \hat{i} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \\ F'y \\ 0 \end{bmatrix} \quad [15]$$

Consider the example data of Table 3 on cows in five HYS from four sires. These data were analyzed by five animal models and the solutions for some effects are given in Table 4. One of the models ignored the sire by HYS association and HYS effects were random in the analysis. The second model treated HYS effects as fixed, and the third model employed Equations [15] where  $L'$  was formed as

$$L' = (1/6) \begin{bmatrix} 2 & 2 & 2 & -3 & -3 \\ 2 & 2 & -3 & 2 & -3 \\ -2 & -2 & -2 & 3 & 3 \\ -3 & 2 & 2 & -3 & 2 \end{bmatrix}$$

with one row per sire. In this example, the rows for sires 1 and 3 are equal except for sign and, thus, the row for sire 3 was omitted in the analysis under this model.

The last two models describe the association between animal genotypes (a) and HYS effects (h) through a covariance such that

$$V \begin{bmatrix} h \\ a \end{bmatrix} = \begin{bmatrix} I\sigma_h^2 & S\sigma_{ah} \\ S'\sigma_{ah} & A\sigma_a^2 \end{bmatrix} \quad [16]$$

Assume for all models that

$$\begin{aligned} \sigma_a^2 &= .25\sigma_y^2, \sigma_h^2 = \sigma_e^2 = .375\sigma_y^2, \\ &\text{and } \sigma_{ah} = .03\sigma_y^2 \\ &\text{(for Model [4]),} \\ \text{or } \sigma_{ah} &= .20\sigma_y^2 \text{ (for Model [5]).} \end{aligned}$$

The matrix  $S$  is equivalent to  $F'Z$  in [15]. The additive genetic values of cows making a record in a HYS are correlated with the HYS effect. The first model is similar to the last two models except that the covariance between  $a$  and  $h$  is assumed to be zero. Henderson was

opposed to the use of a covariance to describe the sire by HYS association problem, but the reasoning was not explained.

The solutions to the five models for HYS and sire effects (expressed as differences from the first HYS and first sire, respectively) differ considerably. As the covariance between  $a$  and  $h$  increased, the range of differences for sires and HYS effects decreased, however, the ranking of sires and HYS effects was unaffected. In the model with HYS effects as fixed and using equations [15] caused sires 2 and 3 to rerank from the first model and caused the differences in HYS effects to expand.

Determining which model is most appropriate for this problem depends on which model best describes the sampling process. The covariance model assumes an average degree of association for all HYS and animals. In practice, the degree of association of animal genotypes with HYS effects could differ across herds and may not be a linear relationship. Use of a covariance model requires the inverse of [16], which precludes the use of a rapid method to invert the relationship matrix.

Treating HYS effects as fixed accounts for the bias in the animal genetic solutions, but this predictor is not necessarily best. Forming an  $L'$  matrix and solving [15] on the other hand, could be very tedious. The amount of bias that occurs in genetic evaluations when this association is ignored has not been quantified.

*Preferential Treatment.* A natural consequence of high prices for semen and embryos of genetically superior individuals is that the offspring of such animals are often preferentially treated (PREF). That is, the offspring are provided with a more favorable environment than progeny of average or below average animals. The detection and recording of PREF animals is impossible without daily surveillance of every herd. When PREF animals have been identified, a suggestion has been to remove their records from the data. Some PREF animals can be viewed as the result of selection on residual effects,  $e$ , in the linear model,

TABLE 4. Solutions for sires and herd-year-season (HYS) in sire by HYS association problem from five different analyses.

Solution	HYS as Random	HYS as Fixed	L/h Used	Correlation of .10	Correlation of .65
HYS2-HYS1	-119.2	-182.5	-192.7	-82.3	-77.1
HYS3-HYS1	-52.5	-82.5	-103.5	-47.5	-44.3
HYS4-HYS1	-148.3	-237.4	-259.1	-152.5	-143.0
HYS5-HYS1	-296.6	-446.6	-448.5	-289.2	-265.9
Sire2-Sire1	-328.0	-300.8	-298.8	-279.1	-270.3
Sire3-Sire1	-337.2	-274.8	-275.8	-314.1	-293.3
Sire4-Sire1	-655.3	-605.9	-606.9	-641.4	-625.5

or at least the residual that an animal would have received is modified by the owner. The resulting MME are given by Henderson (18).

A simpler approach was suggested by Henderson (28), and that was to add PREF groups to the model and assign cows to these groups. A simple means of grouping cows would be on the basis of the semen price of their sire at the time that individual was conceived. All progeny of one bull would not (and should not) be assigned to the same PREF group. Usually the price of semen for a bull changes as the bull progresses from the initial sampling phase to his first proof to his proof based on second crop daughters. To implement this model, semen price records on bulls would need to be accumulated and then merged with production data on their daughters according to calving dates. This would be a very formidable task requiring the cooperation of all AI studs, and the existence of complete historical price lists on all bulls is doubtful.

Analysis with an animal model with PREF groups (much like treatment group effects) would provide estimates of differences in average treatment given to cows in each group adjusted for all other effects in the model. These differences could then be tested for significance. Also, cows that have been preferentially treated would be adjusted for this treatment by the PREF group effect to which they were assigned. The adjustment would not be completely accurate because the adjustment represents the average of all cows in the PREF group. Thus, a cow that receives very favorable treatment may still be overevaluated, but not as much as when PREF groups are not in the model. The fact that we do not know which cows have been preferentially treated will not

help us determine those cows that are still overevaluated.

Creation of large full-sib groups of cows by embryo transfer may necessitate forming PREF groups on the basis of the value of both sire and dam, rather than on only the sire. If prices for semen are not available, then grouping cows on the basis of type evaluations of their sires and dams for evaluation of production traits might be a possibility. A simple strategy would be to use the type classification of the cow herself. However, if there is a genetic relationship between type and production, then the strategy of using type data for sires, dams, or cows should not be used.

Preferential treatment could also be viewed as an association between animal additive genetic effects and residual effects, which could be handled by published procedures (47). Further efforts are needed in this area to define preferential treatment and to postulate an appropriate selection model.

## CONCLUSIONS

This paper has concentrated on contributions to genetic evaluation of animals from C. R. Henderson. To cover every detail of every significant contribution would produce a book (23). The most important contribution was BLUP and the MME. Nearly all animal breeding research today relies on these methods. However, even Henderson questioned whether we should continue to require that predictors be unbiased.

The discovery of a method to obtain the inverse of the additive genetic relationship matrix was equally important and useful. This led to the application of animal models to genetic evaluations and to research into nonadditive

genetic effects. Accounting for inbreeding and selected base animals, and the appropriateness of phantom parent groups with respect to additive genetic and nonadditive genetic relationship matrices need to be solved.

Henderson made numerous contributions to the area of effects of selection on genetic evaluations. This area of research will be important but difficult in the next decade. Estimation of genetic parameters and breeding values from data subject to selection will be a major concern (11, 46). A key will be the accurate descriptions of the sampling process, the rules of selection, and the assumed selection model.

A number of major contributions covering a breadth of topics and having a great impact on the fields of animal breeding and statistics have resulted from the efforts of one man, C. R. Henderson. The applied science of animal breeding will forever be in his debt.

#### REFERENCES

- 1 Chang, H. L., R. L. Fernando and D. Gianola. 1989. Inverse of an additive  $\times$  additive relationship matrix due to sires and maternal grandsires. *J. Dairy Sci.* 72: 3023.
- 2 Ducrocq, V. 1990. Estimation of genetic parameters arising in nonlinear models. *Proc. 4th World Congr. Genet. Appl. Livest. Prod., Edinburgh XIII*:419.
- 3 Gianola, D. 1990. Can BLUP and REML be improved upon? *Proc. 4th World Congr. Genet. Appl. Livest. Prod., Edinburgh XIII*:445.
- 4 Gianola, D., and R. L. Fernando. 1986. Bayesian methods in animal breeding theory. *J. Anim. Sci.* 63: 217.
- 5 Gianola, D., R. L. Fernando, S. Im, and J. L. Foulley. 1989. Likelihood estimation of quantitative genetic parameters when selection occurs: models and problems. *Genome* 31:768.
- 6 Gianola, D., J. L. Foulley, and R. L. Fernando. 1986. Prediction of breeding values when variances are not known. *Genet. Sel. Evol.* 18:485.
- 7 Gianola, D., S. Im, and R. L. Fernando. 1988. Prediction of breeding value under Henderson's selection model: a revisitation. *J. Dairy Sci.* 71:2790.
- 8 Goldberger, A. S. 1962. Best linear unbiased prediction in the generalized linear regression model. *J. Am. Stat. Assoc.* 57:369.
- 9 Henderson, C. R. 1948. Estimation of general, specific, and maternal abilities. Ph.D. Diss., Iowa State Univ., Ames.
- 10 Henderson, C. R. 1949. Estimation of changes in herd environment. *J. Dairy Sci.* 32:706.(Abstr.)
- 11 Henderson, C. R. 1950. Estimation of genetic parameters. *Ann. Math. Stat.* 21:309.(Abstr.)
- 12 Henderson, C. R. 1952. Specific and general combining ability. Ch. 2. Heterosis. Iowa State Coll. Press, Ames.
- 13 Henderson, C. R. 1963. Selection index and expected genetic advance. NAS-NRC 982, Natl. Acad. Sci., Washington, DC.
- 14 Henderson, C. R. 1965. A method for determining if selection will bias estimates. *J. Anim. Sci.* 24: 849.(Abstr.)
- 15 Henderson, C. R. 1973. Sire evaluation and genetic trends. Page 10 in *Proc. Animal Breeding and Genetics Symposium in Honor of Dr. J. L. Lush, ADSA Am. Soc. Anim. Sci., Champaign, IL.*
- 16 Henderson, C. R. 1974. General flexibility of linear model techniques for sire evaluation. *J. Dairy Sci.* 57: 963.
- 17 Henderson, C. R. 1975. Comparison of alternative sire evaluation methods. *J. Anim. Sci.* 41:760.
- 18 Henderson, C. R. 1975. Best linear unbiased estimation and prediction under a selection model. *Biometrics* 31:423.
- 19 Henderson, C. R. 1975. Use of all relatives in intra-herd prediction of breeding values and real producing abilities. *J. Dairy Sci.* 58:1910.
- 20 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.
- 21 Henderson, C. R. 1977. Best linear unbiased prediction of breeding values not in the model for records. *J. Dairy Sci.* 60:783.
- 22 Henderson, C. R. 1982. Best linear unbiased prediction in populations that have undergone selection. *Proc. World Congr. Sheep Beef Cattle Breeding, The Dunmore Press Ltd., N. Z.*
- 23 Henderson, C. R. 1984. Applications of linear models in animal breeding. Univ. Guelph, Guelph, ON, Can.
- 24 Henderson, C. R. 1985. Best linear unbiased prediction using relationship matrices derived from selected base populations. *J. Dairy Sci.* 68:443.
- 25 Henderson, C. R. 1985. Equivalent linear models to reduce computations. *J. Dairy Sci.* 68:2267.
- 26 Henderson, C. R. 1985. Best linear unbiased prediction of nonadditive genetic merits in noninbred populations. *J. Anim. Sci.* 60:111.
- 27 Henderson, C. R. 1990. Statistical methods in animal improvement: historical overview. Page 2 in *Int. Symp. Advances Stat. Methods for Genet. Improve. Livest., Armidale, New South Wales, Aust., Feb. 16-20, 1987.*
- 28 Henderson, C. R. 1990. Accounting for selection and mating biases in genetic evaluations. Page 413 in *Int. Symp. Advances Stat. Methods for Genet. Improve. Livest., Armidale, New South Wales, Aust., Feb. 16-20, 1987.*
- 29 Henderson, C. R. 1988. Exact prediction error variances for full model computed from reduced model. *J. Dairy Sci.* 71:3128.
- 30 Henderson, C. R. 1988. Simple method to compute biases and mean squared errors of linear estimators and predictors in a selection model assuming multivariate normality. *J. Dairy Sci.* 71:3135.
- 31 Henderson, C. R. 1988. A simple method to account for selected base populations. *J. Dairy Sci.* 71:3399.
- 32 Henderson, C. R. 1989. Prediction of merits of potential matings from sire-maternal grandsire models with nonadditive genetic effects. *J. Dairy Sci.* 72:2592.
- 33 Henderson, C. R., O. Kempthorne, S. R. Searle, and

- C. M. vonKrosigk. 1959. The estimation of environmental and genetic trends from records subject to culling. *Biometrics* 15:192.
- 34 Henderson, C. R., and R. L. Quaas. 1976. Multiple trait evaluation using relatives' records. *J. Anim. Sci.* 43:1188.
- 35 Kennedy, B. W., and L. R. Schaeffer. 1989. Genetic evaluation under an animal model when identical genotypes are represented in the population. *J. Anim. Sci.* 67:1946.
- 36 Lush, J. L. 1933. The bull index problem in the light of modern genetics. *J. Dairy Sci.* 16:501.
- 37 McAllister, A. J., J. A. Vesely, T. R. Batra, A. J. Lee, C. Y. Lin, G. L. Roy, J. M. Wauthy, K. A. Winter, and L. A. McClelland. 1990. Genetic changes in protein, milk, and fat yields as a response to selection for protein yield in a closed population of Holsteins. *J. Dairy Sci.* 73:1593.
- 38 Misztal, I. 1990. Restricted maximum likelihood estimation of variance components in animal model using sparse matrix inversion and a supercomputer. *J. Dairy Sci.* 73:163.
- 39 Mood, A. M. 1950. Introduction to the theory of statistics. McGraw-Hill, New York, NY.
- 40 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. R. Soc. Lond. A* 200:1.
- 41 Pollak, E. J., and R. L. Quaas. 1983. Definition of group effects in sire evaluation models. *J. Dairy Sci.* 66:1503.
- 42 Quaas, R. L. 1976. Computing the diagonal elements and inverse of a large numerator relationship matrix. *Biometrics* 32:949.
- 43 Quaas, R. L. 1988. Additive genetic model with groups and relationships. *J. Dairy Sci.* 71:1338.
- 44 Quaas, R. L. 1989. Transformed mixed model equations: a recursive algorithm to eliminate  $A^{-1}$ . *J. Dairy Sci.* 72:1937.
- 45 Quaas, R. L., and E. J. Pollak. 1980. Mixed model methodology for farm and ranch beef cattle testing programs. *J. Anim. Sci.* 51:1277.
- 46 Schaeffer, L. R. 1987. Estimation of variance components under a selection model. *J. Dairy Sci.* 70:661.
- 47 Schaeffer, L. R., and C. R. Henderson. 1983. Best linear unbiased prediction when error vector is correlated with other random vectors in the model. *Genet. Sel. Evol.* 15:395.
- 48 Smith, S. P., and A. Maki-Tanila. 1990. Genotypic covariance matrices and their inverses for models allowing dominance and inbreeding. *Genet. Sel. Evol.* 22:65.
- 49 Tempelman, R. J., and E. B. Burnside. 1990. Additive and nonadditive genetic variation for production traits in Canadian Holsteins. *J. Dairy Sci.* 73:2206.
- 50 Tier, B. 1990. Computing inbreeding coefficients quickly. *Genet. Sel. Evol.* 22:419.
- 51 Van Vleck, L. D. 1982. Derivation of Henderson's method of incorporating AI sire evaluations into intra-herd prediction of breeding values. *J. Dairy Sci.* 65:284.
- 52 Westell, R. A., R. L. Quaas, and L. D. Van Vleck. 1988. Genetic groups in an animal model. *J. Dairy Sci.* 71:1310.