

# Use of Reproductive Technology to Estimate Variances and Predict Effects of Gene Interactions

P. M. VanRADEN

Animal Improvement Programs Laboratory  
Agricultural Research Service, USDA  
Beltsville, MD 20705-2350

T. J. LAWLOR and T. H. SHORT

Holstein Association  
Brattleboro, VT 05302-0808

I. HOESCHELE

Department of Dairy Science  
Virginia Polytechnic Institute and State University  
Blacksburg 24061-0315

## ABSTRACT

Advanced reproductive techniques are creating the large numbers of close relatives needed to study gene interactions. Identical triplets, a set of 26 full sisters, a family of 4215 three-quarter sisters (same sire and maternal grandsire), a family of 76,698 half sisters, and 1.6 million granddaughters of Round Oak Rag Apple Elevation now have lactation records. Similarity of closest relatives might be explained by similar nonadditive as well as additive genetic merit. The 23,015 families of full sisters with mean family size of 3 provide nearly as much information about dominance variation as do the 55,779 families of three-quarter sisters with mean family size of 13; the 79 families of clones provide little information by comparison. Hypothetically, REML analysis of all US Holstein data could provide estimates of dominance and additive  $\times$  additive variance with standard errors approximately 1% of phenotypic variance, but estimates of any higher order interactions would have standard errors  $>10\%$ . The tilde-hat approximation proved to be incompatible with animal models but was used for sire-maternal grandsire analysis of 765,868 first lactation records.

Dominance variance was estimated as 3.5% of phenotypic variance for milk and 3.3% for fat with standard error of 4.2%. With constant data set size, variances are estimated most precisely if family sizes equal 1 plus ratio of within-family to between-family variance. An animal model evaluation including dominance relationships for 581,670 animals was computed, but gene interactions from distant ancestor pairs were ignored. Mating advice and improved additive predictions, especially for clones, could be obtained by including dominance in models.

**(Key words:** reproductive technology, gene interaction, clones, dominance variance)

**Abbreviation key:** ET = embryo transfer, MGS = maternal grandsire, REL = reliability, RIP = record in progress.

## INTRODUCTION

Dairy cattle reproduce less by natural mating and more by AI, embryo transfer (ET), embryo-splitting, and nuclear cloning each year. The large groups of close relatives thus produced share similar main effects of genes (additive effects) and also gene interactions (nonadditive effects). Investigation of gene interactions has been difficult because of low mean nonadditive relationships (3) and lack of appropriate statistical and computational methods. Increasing numbers of identical

---

Received October 1, 1991.

Accepted December 27, 1991.

TABLE 1. Contributions of additive and nonadditive variances to covariances of close relatives.

Relative	Additive	Domi- nance	Additive × additive	Additive × dominance
Identicals	1.0	1.0	1.0	1.0
Full sibs	.5	.25	.25	.125
Parent- progeny	.5	0	.25	0
Three- quarter sibs	.3125	.0625	.0977	.0195
Half sibs	.25	0	.0625	0

animals, full sibs, and three-quarter sibs and the potential for including nonadditive relationships in genetic evaluations now make the study of gene interactions more feasible.

Relationship coefficients for nonadditive genetic effects were established long before corresponding variances ( $\sigma_{ij}^2$ ) could be estimated accurately. Covariances among records of noninbred relatives  $x$  and  $y$  can be calculated by the well-known formula (2)

$$\text{Cov}(x,y) = \sum_i \sum_j a_{xy}^i d_{xy}^j \sigma_{ij}^2, \quad [1]$$

where  $a_{xy}$  is additive relationship of  $x$  and  $y$ , and  $d_{xy}$  is dominance relationship of  $x$  and  $y$ . Summation for  $i$  and  $j$  is from 0 to  $n$  and includes only the combinations where  $1 \leq (i + j) \leq n$ ;  $n$  is the limit for number of genes involved in interactions, and genes are assumed to be unlinked. Thus, probabilities that  $x$  and  $y$  share single genes or pairs of genes at the same locus determine their relationship coefficients for all higher interactions. Examples of these coefficients are in Table 1.

Variances of gene interactions are estimated from covariances among various types of relatives. Recent studies have included families of full sisters (8, 9, 10) produced by ET or families of three-quarter sisters (4) produced by AI. Estimation methods have included REML (8, 9), tilde-hat approximation to REML (4), and noniterative methods (1, 10). Estimates of dominance variance for milk and fat (1, 8) have been inconsistent, have had large standard errors, and sometimes have been outside parameter limits (1). Nonadditive variance esti-

mates were as large as additive variance (4) for cow fertility and from .1 to 4 times additive variance for linear type traits (9, 10). Accurate estimation of nonadditive variances is difficult because proportions of variance shared by relatives may be small and confounded with other genetic or environmental effects.

Genetic evaluations soon could include gene interactions if sufficient variation is present and estimated accurately. Inverses of dominance (5) and additive  $\times$  additive (11) relationship matrices can be constructed rapidly if effects representing interactions of animal pairs are included in matrices. Predictions of additive effects would be more accurate, and predicted interactions would affect mating plans (5, 11). Genetic merits of identical animals can be solved from one equation per clone family instead of per animal, but separate permanent environment equations are required with repeated records (6).

The objectives of this study were 1) to examine numbers and types of relatives available, 2) to compute and compare standard errors of nonadditive variance estimates from various data sets, 3) to estimate dominance variance for milk and fat yields by approximate REML, and 4) to examine computational feasibility of including dominance relationships in genetic evaluation with an animal model.

## MATERIALS AND METHODS

Data files used to produce national genetic evaluations also contain much information regarding nonadditive genetic parameters. Largest family sizes, mean sizes, and number of families present in US Holstein yield data available for January 1991 evaluations are in Table 2. Embryo-splitting and nuclear transfer have been used only recently, and only 79 such clone families had data. Full-sister families in Table 2 were produced only by ET; the largest included 26 full sisters in eight herds in three states. Matings of popular AI bulls to daughters of other popular bulls have produced families of up to 4215 three-quarter sisters. The largest half-sister family includes more than 75,000 daughters of Marshfield Elevation Tony. His sire, Round Oak Rag Apple Elevation, produced the largest quarter-sister family and now has more than 1.6 million granddaughters (Table 3).

TABLE 2. Numbers and types of close relatives in USDA yield data files as of January 1991 and numbers of ideal-sized families that would provide equivalent information.

Family type	Technology used <sup>1</sup>	Number of families	Mean family size	Largest family size	Ideal family size <sup>2</sup>	Balanced number <sup>3</sup>
Identicals	ES	79	2	3	3	49
Full sisters	ET	23,015	3	26	6	6914
Three-quarter sisters	AI	55,779	13	4215	11	33,326
Half sisters	AI	291,587	40	76,698	15	151,268
Quarter sisters	AI	...	...	843,760	63	...

<sup>1</sup>ES = Embryo-splitting; ET = embryo transfer.

<sup>2</sup>Calculated from assumed variances of 25% for additive, 10% for dominance, 5% for additive × additive, and 60% for error.

<sup>3</sup>Number of ideal-sized families required to provide information equivalent to actual unbalanced population.

Three data sets were constructed to investigate 1) presence of gene interactions affecting milk and fat yields, 2) potential to account for gene interactions in animal model evaluations of large populations, and 3) effects of including dominance relationships on additive solutions. To estimate dominance variance, standardized first lactation milk and fat records were obtained from daughters and maternal granddaughters of 1003 popular bulls. These bulls included 879 bulls born from 1970 through 1982 with daughters in more than 250 herds plus 124 older bulls for relationship ties. Edited data were 765,868 records in 285,711 herd-year-seasons and 100,917 filled sire-maternal grandsire (MGS) subclasses; records in progress were excluded. Because herdmates from less popular sires or MGS were excluded, records per herd-year-season were only 2.7, despite 6-mo seasons.

The model of analysis was

$$y_{ijkl} = h_i + s_j + m_k + sm_{jk} + bF_{jk} + e_{ijkl}$$

where  $y_{ijkl}$  is an observation in herd-year-season  $h_i$  from sire  $s_j$ , MGS  $m_k$ , interaction  $sm_{jk}$  with inbreeding coefficient  $F_{jk}$  and regression coefficient  $b$ , and error  $e_{ijkl}$ . Inbreeding was calculated from pedigrees traced only through sire and MGS pathways. Sires and MGS included only the most popular bulls and were treated as fixed to avoid biased estimation of additive variance. Random effects included only  $sm_{jk}$  and  $e_{ijkl}$ ; dominance relationships among  $sm_{jk}$  were included (5). Sire-MGS interactions should have little confound-

ing with genotype × herd interaction because three-quarter sisters usually are located in many different herds. Variance of interactions was estimated by the tilde-hat approach (12).

To explore the feasibility of including dominance relationships in animal models, standardized first lactation records were obtained for 106 clones, 52,040 ET full sisters, and their herdmates. Records with less than 200 d in milk were discarded, and total number of observations was 392,897. The model of analysis was

$$y_{ijk} = h_i + a_j + d_j + bF_j + e_{ijk} \quad [2]$$

where  $y_{ijk}$  is an observation for clone member  $k$  of genotype  $j$ . Variance assumptions in-

TABLE 3. Milk recorded granddaughters of Round Oak Rag Apple Elevation (registration number 1491007) by country and path of descent.

Granddaughter location	Path of descent <sup>1</sup>	Number of granddaughters
US	3425 Elevation sons	802,862
	Elevation daughters	40,898
France	163 Elevation sons	193,102
Canada	150 Elevation sons	180,694
Germany	333 Elevation sons	148,060
The Netherlands	115 Elevation sons	125,181
Italy	156 Elevation sons	119,595
All granddaughters		1,610,392

<sup>1</sup>Only sons with genetic evaluations released on computer media were included.

cluded additive genetic relationships among  $a_j$ , dominance relationships among  $d_j$ , mutually independent  $e_{ijk}$ , and fixed  $h_i$  and  $b$ . Effects of inbreeding on variance of  $d_j$  and covariance of  $a_j$  and  $d_j$  were ignored when the relationship inverses were constructed (5).

To compare solutions with and without dominance relationships, a smaller data set for the trait stature was examined. Clones and their herdmates were included for a total of 6750 records. By chance, 537 full-sister families were included with 7 full sisters in the largest family. These data were analyzed with the model in Equation [2] and with a model that was identical, except that  $d_j$  was removed and dominance relationships were ignored. Additive and dominance variances were assigned arbitrary values of 27 and 21% of phenotypic variance, respectively.

Estimates of genetic variances are most precise if data contain large numbers of several types of close relatives. The number of family types must equal or exceed the number of genetic variances to estimate. Thus, additive and dominance variance might be estimated from just full and half sibs (8, 9) whereas estimation of additive  $\times$  additive variation as a third genetic component would require three types of relatives, such as three-quarter, half, and quarter sibs (1, 4). Quarter sibs, also known as half cousins, have just one grandparent in common.

Analyses using REML or approximate REML that include additive and nonadditive relationship matrices among animals should allow all types of relatives to contribute to estimation of genetic parameters. Although exact standard error estimates generally cannot be computed, inexpensive approximations can be obtained from variances of simple quadratic forms.

**Variances of Quadratic Forms**

Variation within and among families can be summarized by simple sums of squares, especially with balanced data in which family size is constant. Variances of these sums of squares also can be computed easily by known rules to provide standard errors of parameter estimates. Information within and among families of clones, full sibs, half sibs, and parent-offspring pairs can be summarized by computing the

following two quadratic forms for each relative type.

If  $f$  families of a particular type each have  $m$  members, variation within families can be summarized as

$$q_0 = \left[ \sum_i \sum_j y_{ij}^2 - \sum_i \left( \sum_j y_{ij} \right)^2 / m \right] [fm(m - 1)].$$

where  $i$  denotes family,  $j$  denotes member within family, and  $y_{ij}$  is an observation expressed as a deviation from population or herd-year-season mean so that expected value of  $y_{ij}$  is 0. Variation among families can be summarized as

$$q_1 = \left[ \sum_i \left( \sum_j y_{ij} \right)^2 - \sum_i \sum_j y_{ij}^2 \right] [fm(m - 1)].$$

Quadratic  $q_1$  differs from the usual among-family sum of squares,  $\sum_i \left( \sum_j y_{ij} \right)^2$  because expected contribution of within-family variance has been removed. By subtracting  $\sum_i \sum_j y_{ij}^2$ , among-family variance is estimated directly.

The two quadratics can be written in matrix notation as

$$q_0 = \mathbf{y}'\mathbf{Q}_0\mathbf{y};$$

$$q_1 = \mathbf{y}'\mathbf{Q}_1\mathbf{y}$$

with

$$\mathbf{Q}_0 = \left( \mathbf{I}_m - \mathbf{1}_m\mathbf{1}'_m/m \right) \otimes \mathbf{I}_f [f(m - 1)];$$

$$\mathbf{Q}_1 = \left( \mathbf{1}_m\mathbf{1}'_m - \mathbf{I}_m \right) \otimes \mathbf{I}_f [fm(m - 1)]$$

where subscripts  $m$  and  $f$  denote dimensions of vectors or matrices,  $\mathbf{I}$  is an identity matrix,  $\mathbf{1}$  is a column vector of 1's, and  $\otimes$  denotes Kronecker product. Vector  $\mathbf{y}$  contains observations  $y_{ij}$  for all families of one type and has mean and variance

$$E(\mathbf{y}) = \mathbf{0};$$

$$\text{Var}(\mathbf{y}) = \left( \mathbf{I}_m\sigma_w^2 + \mathbf{1}_m\mathbf{1}'_m\sigma_f^2 \right) \otimes \mathbf{I}_f$$

where  $\sigma_f^2$  denotes genetic covariance among family members as obtained from Equation [1], and  $\sigma_w^2$  denotes the phenotypic variance within such families. Covariance structure among actual families is more complicated because, for example, sires of half-sister families often are related rather than unrelated, as assumed here, and because family sizes are unbalanced.

Mean and variance of quadratics for normally distributed data are obtained from matrix identities  $E(\mathbf{y}'\mathbf{Q}\mathbf{y}) = \text{tr}[\mathbf{Q}\text{Var}(\mathbf{y})]$  and  $\text{Var}(\mathbf{y}'\mathbf{Q}\mathbf{y}) = 2\text{tr}[\mathbf{Q}\text{Var}(\mathbf{y})\mathbf{Q}\text{Var}(\mathbf{y})]$ . Expectations and variances of  $q_0$  and  $q_1$  require evaluating the matrix products

$$\mathbf{Q}_0\text{Var}(\mathbf{y}) = (\mathbf{I}_m\sigma_w^2 - \mathbf{I}_m\mathbf{I}_m'\sigma_w^2) \otimes \mathbf{I}_f[f(m-1)];$$

$$\mathbf{Q}_1\text{Var}(\mathbf{y}) = \left\{ \mathbf{I}_m\mathbf{I}_m' \left[ \sigma_w^2 + (m-1)\sigma_f^2 \right] - \mathbf{I}_m\sigma_w^2 \right\} \otimes \mathbf{I}_f[fm(m-1)].$$

Evaluation of the trace function and further algebra gives

$$E(q_0) = \sigma_w^2;$$

$$\text{Var}(q_0) = 2\sigma_w^4/[f(m-1)]$$

and

$$E(q_1) = \sigma_f^2;$$

$$\text{Var}(q_1) = 2 \left[ m\sigma_f^4 + \sigma_w^4/(m-1) + 2\sigma_f^2\sigma_w^2 \right] / (fm).$$

Because covariance of  $\mathbf{y}'\mathbf{Q}_0\mathbf{y}$  and  $\mathbf{y}'\mathbf{Q}_1\mathbf{y}$  is  $2\text{tr}[\mathbf{Q}_0\text{Var}(\mathbf{y})\mathbf{Q}_1\text{Var}(\mathbf{y})]$ ,

$$\text{Cov}(q_0, q_1) = -2\sigma_w^4/[fm(m-1)].$$

#### Estimation of Genetic Parameters

Variation among and within each family type is expected to contain the fractions of genetic variance in Table 1. Let  $\mathbf{v}$  represent the vector of additive, dominance, additive  $\times$  additive, and environmental variance components to estimate, and let  $\mathbf{H}$  represent the fractions of each variance expected within and among each

family type. If vector  $\mathbf{q}$  includes quadratics  $q_0$  and  $q_1$  for each family type, expectation and variance of  $\mathbf{q}$  are

$$\begin{aligned} E(\mathbf{q}) &= \mathbf{H}\mathbf{v}; \\ \text{Var}(\mathbf{q}) &= \mathbf{V}_q. \end{aligned}$$

Diagonals of  $\mathbf{V}_q$  equal respective  $\text{Var}(q_0)$  and  $\text{Var}(q_1)$ . Quadratics for different family types were assumed to be uncorrelated, even though the same animals may be included in more than one family type. Thus, off-diagonals of  $\mathbf{V}_q$  were set equal to 0 except for  $q_0$  and  $q_1$  of the same family type.

For simplicity, the number of quadratics included in  $\mathbf{q}$  may be limited to only the number of variances to be estimated so that  $\mathbf{H}$  is square and nonsingular. Variance estimates then can be computed as

$$\hat{\mathbf{v}} = \mathbf{H}^{-1}\mathbf{q}.$$

With more family types and  $\mathbf{q}$  longer than  $\mathbf{v}$ , generalized least squares can be used. Then,  $\mathbf{q}$  is treated as an ordinary vector of data and estimates are obtained as

$$\hat{\mathbf{v}} = (\mathbf{H}'\mathbf{V}_q^{-1}\mathbf{H})^{-1}\mathbf{H}'\mathbf{V}_q^{-1}\mathbf{q}.$$

For these generalized least squares estimates, standard errors are obtained from

$$\text{Var}(\hat{\mathbf{v}}) = (\mathbf{H}'\mathbf{V}_q^{-1}\mathbf{H})^{-1}.$$

Even for variances estimated with other methods such as REML or approximate REML, standard errors computed with this formula may serve as good approximations. As usual, constraints on parameter space are ignored in the standard error formula. Thus, computed standard errors may be too large, and confidence ranges could include negative values. Elements of  $\mathbf{V}_q$  can be calculated before estimates are obtained from prior or starting values of variance components or after analysis from converged estimates.

#### Optimal Family Size

Families of intermediate size may provide maximum information regarding variance components. Suppose that the population size

(n) is constant, and  $n = fm$ . At the extremes, no information is provided either by  $n$  families containing only one member each ( $m = 1$ ) or by only one extremely large family ( $f = 1$ ) with  $n$  members. The family size that provides maximum information about  $\sigma_f^2$  is 1 plus ratio of within-family to between-family variance (7), which can be derived from the partial derivative

$$\partial \text{Var}(q_1) / \partial m = (2/n) \left[ \sigma_f^4 - \sigma_w^4 / (m - 1)^2 \right].$$

If  $\partial \text{Var}(q_1) / \partial m$  is set equal to 0, then

$$(m - 1)^2 = \sigma_w^4 / \sigma_f^4.$$

Finally ideal family size ( $m$ ) is obtained:

$$m = \sigma_w^2 / \sigma_f^2 + 1.$$

Many different family sizes may be present in unbalanced data. Quadratics can be modified to accommodate unequal family size, or actual data structure can be converted to an equivalent, balanced design. Family size in the hypothetical balanced design can be chosen as the ideal size or, perhaps, as the median family size to represent actual data more closely. For each actual family of size  $m$ , equivalent number ( $x$ ) of ideal-sized or median-sized families of size  $s$  is calculated as the ratio of  $\text{Var}(q_1)$  evaluated at  $s$  and  $m$ , or

$$x = \frac{m \left[ s\sigma_f^4 + \sigma_w^4 / (s - 1) + 2\sigma_f^2\sigma_w^2 \right]}{s \left[ m\sigma_f^4 + \sigma_w^4 / (m - 1) + 2\sigma_f^2\sigma_w^2 \right]}.$$

The  $x$  balanced families then provide a standard error for estimating  $\sigma_f^2$  equivalent to that of the actual family. Total number of balanced families is determined by summing  $x$  across all actual families.

#### Tilde-Hat with Animal Model

Approximate REML estimates can be obtained from many data sets for which exact REML cannot be computed. Unfortunately, the tilde-hat approximation (12) appears to be ill suited for use with animal models. Consider a

simple model that has only additive relationships, each animal with one record ( $\mathbf{Z} = \mathbf{I}$ ), and no fixed effects to estimate. With the ratio  $\sigma_e^2 / \sigma_a^2$  denoted as  $k$ , "tilde" and "hat" solutions for the vector of additive effects  $\mathbf{u}$  are  $\tilde{\mathbf{u}} = \mathbf{y} / (1 + k)$  and  $\hat{\mathbf{u}} = (\mathbf{I} + \mathbf{A}^{-1}k)^{-1}\mathbf{y}$ .

Variances are estimated by setting the quadratic  $\tilde{\mathbf{u}}'\mathbf{A}^{-1}\tilde{\mathbf{u}}$  equal to its pseudoexpectation,  $\sigma_a^2 n / (1 + k)$ , and the quadratic  $\mathbf{y}'\mathbf{y} - \mathbf{y}'\hat{\mathbf{u}}$  equal to its pseudoexpectation,  $\sigma_e^2 n$ , where  $n$  denotes number of animals. In each round of iteration, new variance estimates would be

$$\hat{\sigma}_a^2 = \mathbf{y}'\mathbf{A}^{-1}(\mathbf{I} + \mathbf{A}^{-1}k)^{-1}\mathbf{y} / n;$$

$$\hat{\sigma}_e^2 = [\mathbf{y}'\mathbf{y} - \mathbf{y}'(\mathbf{I} + \mathbf{A}^{-1}k)^{-1}\mathbf{y}] / n.$$

Multiplication by  $k$  of numerator and denominator followed by addition and subtraction of  $\mathbf{I}$  allows the expression for  $\hat{\sigma}_a^2$  to be rewritten:

$$\hat{\sigma}_a^2 = \mathbf{y}'(\mathbf{I} + \mathbf{A}^{-1}k - \mathbf{I})(\mathbf{I} + \mathbf{A}^{-1}k)^{-1}\mathbf{y} / (nk)$$

$$= [\mathbf{y}'\mathbf{y} - \mathbf{y}'(\mathbf{I} + \mathbf{A}^{-1}k)^{-1}\mathbf{y}] / (nk).$$

The new ratio  $\hat{\sigma}_e^2 / \hat{\sigma}_a^2$  always equals  $k$ , the previously supplied ratio. The algorithm makes no progress because the two quadratics are identical functions of the data divided by different constants. Similar algebra shows that this problem also occurs for animal models with fixed effects. The algorithm loses all information from relatives because each "tilde" solution contains the animal's own record only. The tilde-hat approach works with other models because data from descendants are accumulated directly into the right-hand sides for sires, dams, or MGS.

## RESULTS

Dominance variance was estimated to be 3.5% of phenotypic variance for milk and 3.3% for fat with standard errors of  $\pm 4.2\%$  in an analysis of 765,868 records. Variance of sire-MGS interactions was only .22% of phenotypic variance for milk and .21% for fat. Corresponding estimates were .19% for both traits if relationships among interactions were ignored. Main effects of sire and MGS were

TABLE 4. Previous estimates of dominance and additive  $\times$  additive variance as percentage of phenotypic variance and standard errors computed from reported estimates and family numbers, sizes, and types.

Authors	Year	Reference	Trait	Variance estimate					
				Additive		Dominance		Additive $\times$ additive	
				%	SE	%	SE	%	SE
Allaire and Henderson	1965	(1)	Milk Fat <sup>1</sup>	11	6	50	26	16	31
				23	10	-18	35	24	49
Tempelman and Burnside	1990	(8) <sup>2</sup>	Milk Fat	40	2	6	6	...	...
				32	2	24	6	...	...
Tempelman and Burnside	1990	(9) <sup>2</sup>	Type	15	2	15	6	...	...
Hoeschele	1991	(4)	Days open	2	1	2	4	1	4

<sup>1</sup>Standard error calculations assumed dominance variance of 0 instead of -18.

<sup>2</sup>Reported standard errors ranged from 2 to 3% for additive variance and from 8 to 10% for dominance variance.

assumed to contribute 7.8% [(0.25 + .0625)25%] of phenotypic variance, even though these terms were considered to be fixed in the model. Interactions were expected to contain 1/16 of dominance variance; contributions of additive  $\times$  additive and higher order variances were ignored.

Sire and MGS evaluations were similar to published values because of large numbers of offspring included. Mean inbreeding of cows was only .67%, because pedigrees ended with the first ancestors born before 1970 and were not traced through maternal granddams. Regressions on inbreeding were -26 kg/1% for milk and -.9 kg/1% for fat.

Estimates from previous studies of nonadditive genetic variation are in Table 4; approximate standard errors computed from numbers and types of relatives are included. Only Tempelman and Burnside (8, 9) published standard errors of their estimates. Although approximate standard errors from this study are slightly smaller than their published standard errors, they agree well, considering the combined effects of the approximated data structure and generalized least squares approach used.

Table 5 shows hypothetical standard errors that might result from analysis of all US Holstein data. Models with fewer nonadditive effects produce smaller standard errors of variance estimates, but these estimates may be biased if the omitted effects have positive variance. Dominance and additive  $\times$  additive vari-

ances could have standard errors that are 1% of phenotypic variance, but little information is available regarding higher order interactions, such as additive  $\times$  dominance. Attempts to estimate higher order parameters may greatly increase standard errors for other components.

Table 6 shows hypothetical standard errors from including only certain types of relatives from the Holstein population. Dominance estimates with standard errors of 3% of phenotypic variance might be produced using data from either three-quarter or full sibs. Clone families have all of their nonadditive effects in common but currently are too few in number to add much information about nonadditive variances beyond that provided by the more numerous full and three-quarter sibs.

TABLE 5. Standard errors of additive (A), dominance (D), additive  $\times$  additive (AA), and additive  $\times$  dominance (AD) variance estimates using all US Holstein data to estimate variances from progressively more complete models.

Model effects	Standard error <sup>1</sup>				Error
	A	D	AA	AD	
A	.12				.1
A, D	.12	1.4			1.4
A, D, AA	.35	1.4	.86		1.5
A, D, AA, AD	.36	4.7	.87	10.5	6.1
Assumed variances <sup>1</sup>	25	10	5	2	58+

<sup>1</sup>Expressed as percentage of phenotypic variance.

TABLE 6. Standard errors of additive (A), dominance (D), and additive × additive (AA) variance estimates from hypothetical analyses including selected family types from US Holstein data.

Clone	Family type					Standard error <sup>1</sup>		
	Full sister	Parent-progeny	Three-quarter sister	Half sister	Quarter sister	A	D	AA
				X	X	X	1.0	2.7
	X			X	X	1.0	3.0	4.3
	X	X	X	X	X	.3	1.4	.9
X	X	X	X	X	X	.3	1.4	.9
Assumed variances <sup>1</sup>						25	10	5

<sup>1</sup>Expressed as percentage of phenotypic variance.

Correlations of genetic variance estimates are shown in Table 7 for data on days open reported by Hoeschele (4). Strong negative correlations for this and for other data sets imply that when the additive × additive variance estimate is higher than its true value, additive and dominance estimates are lower than their true values. Dominance estimates were positively correlated with additive in this model, but negatively correlated in models that omitted additive × additive variance. Knowledge of these correlations may help to explain unusual estimates and to predict how a deletion or bias of one component may affect estimates of others. Covariances of parameter estimates are computed automatically in the generalized least squares approach.

Animal model analysis of Holstein clone and full-sister data required just over 2 million equations and a total of 15 million nonzero coefficients in the additive and dominance inverses (Table 8). Some dominance relationships had to be ignored because ancestor interactions would not all fit in the available 48 Mb

TABLE 7. Correlations among estimates of additive, dominance, and additive × additive variances from data of Hoeschele (4).

Variance component	Additive	Dominance	Additive × additive	Error
Additive	1.00	.39	-.97	.37
Dominance	.39	1.00	-.47	-.68
Additive × additive	-.97	-.47	1.00	-.32
Error	.37	-.68	-.32	1.00

of memory. To reduce memory needs, maternal granddams without records were treated as unknown, and interactions among ancestors more than two generations removed from an animal with data were discarded.

About three times as many equations and nonzero coefficients were generated for dominance compared with additive effects. Except for clones, each animal and ancestor had individual additive and dominance solutions. Dominance equations also were solved for 259,702 filled sire-dam subclasses and 559,641 ancestor subclasses needed to construct the

TABLE 8. Equations and nonzero coefficients required for an animal model that includes dominance relationships for 392,897 cows with records.

Equation type	Number
Regression on inbreeding	1
Herd-year-seasons	27,113
Additive effects	
Genotypes with records	392,844
Ancestor genotypes	188,773
Dominance effects	
Genotypes with records	392,844
Filled sire-dam subclasses	259,702
Ancestor genotypes	188,773
Ancestor sire-dam subclasses <sup>1</sup>	559,641
Total equations	2,009,691
Nonzero coefficients	
Additive inverse	
Unsummed	3,555,563
Summed	2,681,360
Dominance inverse	
Unsummed	11,760,031
Summed	7,474,689

<sup>1</sup>Remote ancestor subclasses were excluded.



TABLE 9. Genetic evaluations of clones with and without dominance in the model for stature data.

Clone members	Additive	Dominance <sup>1</sup>	Additive only
2	+6.90	+5.31	+8.38
2	+5.23	+2.54	+5.71
2	+6.9	+2.81	+1.49
2	-2.39	-5.25	-3.43
3	-5.39	-5.45	-6.85
2	-1.28	-5.67	-2.22

<sup>1</sup>Clone families are those with highest and lowest dominance solutions.

dominance inverse. Dominance relationship matrices have fewer nonzero elements than do additive relationship matrices, but their inverses often have more nonzero coefficients. Reasons are the addition of subclass effects and that submatrices of maximum size  $9 \times 9$  rather than  $3 \times 3$  are needed to construct dominance inverses (5).

The total time required to create and solve the animal model equations was 40 clock h on a Sun workstation (Sun Microsystems, Inc., Mountain View, CA). The large hash tables constructed to process animals and animal interactions nearly filled the available memory. Convergence of solutions to three significant digits required 150 iterations with successive overrelaxation and a relaxation factor of 1.5, the optimal factor as determined from a much smaller data set.

Additive genetic effects of clones were overestimated if assumed dominance covariances were ignored in the stature data set. Clone families with highest and lowest predicted dominance effects had substantial changes in predicted additive merit as shown in Table 9. Dominance effects, which are excluded from the model, may not simply fade into the error term but instead may contribute to estimates of additive effects. Differences in Table 9 are large because of assumed dominance variance of 21% but demonstrate the potential advantage of including dominance relationships for data containing clones.

### CONCLUSIONS

Reproductive technologies that include AI, ET, and embryo-splitting are producing fami-

lies of three-quarter sibs, full sibs, and clones of increasing size and number in dairy cattle. Members of such families share common gene interactions, which often have been ignored in selection and mating programs. Animal models that include a regression on inbreeding and additive, dominance, and additive  $\times$  additive relationships now can account for all two-gene interactions shared by these relatives. All types of relatives would contribute to estimates of dominance and additive  $\times$  additive variance if computed by REML.

Approximate standard errors of variance components can be obtained inexpensively from sizes and numbers of each family type. Information from sums of squares within and among these families can be combined in a generalized least squares approach to approximate information extracted by REML or approximate REML. The tilde-hat approximation (12) was incompatible with the animal model. Smallest standard errors of variance estimates are obtained if all family sizes equal 1 plus ratio of within-family to between-family variance.

Dominance and additive  $\times$  additive variances could be estimated with standard error of only 1% of phenotypic variance if all US Holstein data were included. Computationally affordable analyses can yield standard errors of 2 to 10%. Any estimates of higher order interactions, such as additive  $\times$  dominance, would have standard errors greater than 10%, even from all data. An analysis of 765,868 milk and fat records representing 1003 sires and MGS estimated dominance variance as  $3 \pm 4\%$  of phenotypic variance with no additive  $\times$  additive variance assumed.

An animal model including dominance relationships among 106 clones, 52,040 ET full sisters, 340,751 herdmates, and 188,773 ancestors was computationally affordable, although some remote relationships were excluded. Analysis of a similar but smaller data set showed that additive genetic merits of clones may be inflated if dominance relationships are ignored.

### ACKNOWLEDGMENTS

Financial support from Eastern AI Cooperative and the National Association of Animal Breeders to I. Hoeschele is acknowledged

gratefully. The cooperation of the US dairy industry in supplying yield and pedigree data through the National Cooperative Dairy Herd Improvement Program is appreciated. Both J. Dekkers and R. Tempelman provided helpful comments and corrections to the manuscript.

#### REFERENCES

- 1 Allaire, F. R., and C. R. Henderson. 1965. Specific combining abilities among dairy sires. *J. Dairy Sci.* 48:1096.
- 2 Cockerham, C. C. 1954. An extension of the concept of partitioning hereditary variance for analysis of variance when epistasis is present. *Genetics* 39:859.
- 3 Freeman, A. E., and C. R. Henderson. 1959. Genetic structure of dairy cattle herds in terms of additive and dominance relationships. *J. Dairy Sci.* 42:621.
- 4 Hoeschele, I. 1991. Additive and nonadditive genetic variance in female fertility of Holsteins. *J. Dairy Sci.* 74:1743.
- 5 Hoeschele, I., and P. M. VanRaden. 1991. Rapid inversion of dominance relationship matrices for noninbred populations by including sire by dam subclass effects. *J. Dairy Sci.* 74:557.
- 6 Kennedy, B. W., and L. R. Schaeffer. 1990. Reproductive technology and genetic evaluation. Page 507 in *Advances in Statistical Methods for Genetic Improvement of Livestock*. D. Gianola and K. Hammond, ed. Springer-Verlag, Heidelberg, Germany.
- 7 Robertson, A. 1959. Experimental design in the evaluation of genetic parameters. *Biometrics* 15:219.
- 8 Tempelman, R. J., and E. B. Burnside. 1990. Additive and nonadditive genetic variation for production traits in Canadian Holsteins. *J. Dairy Sci.* 73:2206.
- 9 Tempelman, R. J., and E. B. Burnside. 1990. Additive and nonadditive genetic variation for conformation traits in Canadian Holsteins. *J. Dairy Sci.* 73:2214.
- 10 Thomas, C. L., W. E. Vinson, R. E. Pearson, and H. D. Norman. 1985. Components of genetic variance and covariance for linear type traits in Jersey cattle. *J. Dairy Sci.* 68:2989.
- 11 VanRaden, P. M., and I. Hoeschele. 1991. Rapid inversion of additive by additive relationship matrices by including sire-dam combination effects. *J. Dairy Sci.* 74:570.
- 12 VanRaden, P. M., and Y. C. Jung. 1988. A general purpose approximation to restricted maximum likelihood: the tilde-hat approach. *J. Dairy Sci.* 71:187.