INFORMATION TO USERS

This material was produced from a microfilm copy of the original document. While the most advanced technological means to photograph and reproduce this document have been used, the quality is heavily dependent upon the quality of the original submitted.

The following explanation of techniques is provided to help you understand markings or patterns which may appear on this reproduction.

- The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting thru an image and duplicating adjacent pages to insure you complete continuity.
- 2. When an image on the film is obliterated with a large round black mark, it is an indication that the photographer suspected that the copy may have moved during exposure and thus cause a blurred image. You will find a good image of the page in the adjacent frame.
- 3. When a map, drawing or chart, etc., was part of the material being photographed the photographer followed a definite method in "sectioning" the material. It is customary to begin photoing at the upper left hand corner of a large sheet and to continue photoing from left to right in equal sections with a small overlap. If necessary, sectioning is continued again beginning below the first row and continuing on until complete.
- 4. The majority of users indicate that the textual content is of greatest value, however, a somewhat higher quality reproduction could be made from "photographs" if essential to the understanding of the dissertation. Silver prints of "photographs" may be ordered at additional charge by writing the Order Department, giving the catalog number, title, author and specific pages you wish reproduced.
- 5. PLEASE NOTE: Some pages may have indistinct print. Filmed as received.

Xerox University Microfilms 300 North Zeeb Road Ann Arbor, Michigan 48106

ی و رو د این		\$
	75-3319	1.5
McGILLIARD, Michael Lon, 1947- PREDICTING DAUGHTER MILK PRODUCTION DAM INDEX.	i From	
Iowa State University, Ph.D., 1974 Agriculture, Animal Culture		

.

Xerox University Microfilms, Ann Arbor, Michigan 48106

. . . .

.....

.

THIS DISSERTATION HAS BEEN MICROFILMED EXACTLY AS RECEIVED.

• -

Predicting daughter milk

production from dam index

'by

Michael Lon McGilliard

A Dissertation Submitted to the Graduate Faculty in Partial Fulfillment of The Requirements for the Degree of

DOCTOR OF PHILOSOFHY

Department: Animal Science Major: Animal Breeding

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

For the Major Department

Signature was redacted for privacy.

For the Graduate College

Iowa State University Ames, Iowa

TABLE OF CONTENTS

.

	Page
INTRODUCTION	1
REVIEW OF LITERATURE	3
ESTIMATED AVERAGE TRANSMITTING ABILITY	11
HISTORY OF EATA	30
DATA	32
METHODS	34
RESULTS AND DISCUSSION	40
SUMMARY	106
REFERENCES	109
ACKNOWLEDGEMENTS	112

LIST OF TABLES

Tabl	e	Page
1.	Correlations between son's progeny test and breeding value of dam	7
2.	Regression of son's progeny test on various relatives	7
3.	Computation of an EATA	18
4.	Example of large $(\Sigma b_i - \Sigma b'_i)$	28
5.	Average numbers of relatives, records, and associated weights for EATA	38
6.	Numbers of animals and yield means	41
7.	Correlations between dam EATA and daughter yield	41
8.	Regressions of daughter yield on dam EATA	43
9.	Numbers of animals and yield means for EATA segments	49
10.	Correlations between dam EATA and daughter first lactation for EATA segments	54
11.	Regressions of daughter first lactation on dam EATA within EATA segments	56
12.	Information on low sixty and high sixty dam EATA's	59
13.	Yield means for sixty lowest EATA's, ten classes of six daughters each	61
14.	Yield means for sixty highest EATA's, ten classes of six daughters each	63
15.	Yield characteristics of first lactation daughters and dams with varying numbers of dam records	65
16.	Numbers of daughter lactations available for each combina- tion of relatives of the dam	68
17.	Numbers of animals and average milk yield for specific combinations of EATA relatives	70
18.	Mean numbers of animals and records in EATA's of specific relative combinations	7 4

LIST OF TABLES (continued)

Table	e	Page
19.	Partial regression coefficients to predict daughter first lactation deviated milk yield from relatives in her dam's EATA	76
20.	Standard partial regression coefficients from various combinations of relatives	82
21.	Multiple correlations between daughter first lactation deviation yield and combinations of dam relatives	. 85
22.	Variances and standard deviations of numbers of animals and records in 3414 EATA's with all relative groups	86
23.	Calculated EATA's and R 's based on mean numbers of relatives, records, and production from various combina- tions of relatives	89
24.	Standard deviations of average deviated production of relatives in EATA for combinations of relatives in EATA	92
25.	Correlations b etw een daughter first lactation deviated yield and average deviations of relatives in her dam's EATA	94
26.	Correlations between average deviated yield of relatives in EATA	98
27.	Fractions of daughters with first lactation ME yield greater than herdmates	104

LIST OF FIGURES

Fig	ure	Page
1.	Equations for computing weights for EATA	17
2.	Linear regression of daughter first lactation deviation on dam EATA in relation to expected (E)	45
3.	Quadratic regression of daughter first lactation deviation on dam EATA in relation to expected (E)	46
4.	Cubic regression of daughter first lactation deviation on dam EATA in relation to expected (E)	47
5.	Cubic regression of daughter ME production on dam EATA in relation to expected (E)	48
ó.	Distribution of daughters by dam EATA	51
7.	Expected and observed daughter production for EATA segments	53
8.	Quadratic regression of daughter production on dam EATA in high and low EATA segments in relation to expected (E)	57

ł

•

V

INTRODUCTION

Genetic gain per year for milk production in a dairy population depends on intensity of selection, accuracy of selection, genetic differences in the population, and generation interval. Accuracy, the correlation between phenotype and genotype, increases as more information is obtained on individuals, usually causing an increase in generation interval. The genetic gain from improved accuracy of selection is often suppressed by a longer generation interval. The goal is to increase accuracy without lengthening generation interval, or to shorten generation interval without decreasing accuracy. To accomplish this we need a reliable estimate of an animal's genetic worth for milk production early in its life. The early practical time to estimate the worth of an animal is before breeding age. This dictates use of a breeding value estimated from the individual's ancestors and collateral relatives.

Assertions that sires contribute 69% (Skjervold, 1963) to 61% (Robertson and Rendel, 1950) of potential genetic progress in milk production versus 31% to 39% for dams has lead to much research evaluating methods of estimating genetic worth of dairy bulls. Less research has been devoted to females. These efforts are not disjoint, however, as genetically superior females are identified more accurately, the more likely their sons will be identified as potentially superior sires. Selection for dams alone will contribute more than 50% of maximum genetic progress to the population (Lush, 1960). Regardless of how many offspring a bull sires they each receive half their genes from their dams. It is important dams of sires be as accurately evaluated as possible to save

time and expense of maintaining sires through progeny tests by mistake.

Since a bull does not give milk, a progeny test (Predicted Difference) measures his genetic worth for milk production. Most studies have dealt with the worth of his breeding value based on pedigree information to predict his future PD. These studies are limited by small numbers of bulls available who receive PD's compared with numbers of bulls born. This is due to the impracticality of progeny testing most bulls. Efforts to correlate pedigree estimates of breeding value with milk production in bulls are complicated by an additional gene segregation when they are evaluated through their daughters.

With females it is possible to study the accuracy and precision by which pedigree estimates of breeding value predict future offspring milk production. We can evaluate the predicted genotype through its performance rather than sampling through its offspring performance as necessary for bulls. Many more cows than bulls are available for study, although bulls' breeding values can be more accurately determined.

The increasing popularity of choosing a specific bull to mate to a specific cow to produce a future sire magnifies the importance of accurately appraising the genetic worth of those parents. This appraisal is a pedigree estimate of the breeding value of the offspring.

With much work published on evaluations of the male side of pedigrees this study was designed to evaluate the usefulness of estimated breeding values of dams to predict milk production of their future daughters where dam and daughter are in the same herd.

2a

The following definitions may be useful.

Genetic. Associated with genes.

Effect. Units added or deleted by a variable.

Genic. Associated with the additive effects of genes.

Phenotype. Measured performance (kilograms of milk).

Genotype. Additive contributions of genes to animal's phenotype. Genotype is usually defined as the collection of genes an animal possesses.

Breeding Value. Additive effect of all genes which influence a character.

Heritability. Fraction of differences between records which is genetic.

Repeatability. Fraction of differences between records which is genetic and permanently environmental. Gorrelation between records of a cow.

EATA. Estimated Average Transmitting Ability.

PD. Predicted Difference.

REVIEW OF LITERATURE

Empirical evaluations of selection indexes for milk production in dairy females are scarce. Deaton and McGilliard (1965) calculated from 7638 first lactation Holstein records variances and covariances between the five relatives used in the current EATA, cows, dams, daughters, paternal and maternal sisters. They computed the covariances between a cow's genotype and each relative by multiplying the genic relationship by the observed heritability (.25) and variance of individual records. These variances and covariances were assembled into five equations analogous to the theoretical EATA equations. Various numbers of relatives were substituted into the equations, each yielding a solution of five partial regression coefficients which maximized the correlation of index with cow's genotype. From multiple correlations they concluded dams and maternal sisters contribute little accuracy to estimates of a cow's genic value. Accuracy of indexes for cows without records was low compared with indexes for cows with records. Relatives further back in pedigrees were more useful where cows had no records. Correlation of cow index with unselected first lactation daughter production was .17 where the daughter record was not in the index. Correlation of cow first lactation with daughter first lactation was .14, indicating the index held a 19% advantage in genetic gain for production.

Flock (1964) examined records from 11,630 registered Holstein cows and found correlations between relatives, including cow, dam, maternal sisters, paternal sisters, daughters, and paternal sisters of daughters, 10% smaller than predicted. Exceptions were cow and dam, paternal and mater-

nal sisters, and daughters and paternal sisters of daughters. For a sample of 3518 cows he concluded from correlations between indexes and cows' true breeding values daughters and maternal half sisters contributed little to the accuracy of the index. Using only cows' own records in the index was 91% as accurate for estimating true breeding values as using all six relative groups in the index.

Syrstad (1971) indexed 136,242 Norwegian Red cows using records of the cows and their dams and paternal sisters. The mean index was 102 with a standard deviation of 4.5. He recommended saving bull calves from cows with indexes larger than 111. These composed 3% of his sample. Cows with fewer records averaged higher index values.

Deaton and McGilliard (1964) also studied the accuracy of predicting first lactation deviated daughter production from dam's deviation with Guernseys and Holsteins. Dam's first lactation was a more accurate predictor than second or third lactations or any average. First records were less variable than second or third records. Repeatability of .52 indicated records beyond the first were of little value.

VanVleck and Bradford (1966) found heritability of first lactation Holstein production to be larger than heritability of second or third lactations. They constructed indexes to estimate a cow's first record breeding value from combinations of her own records and records of her dam where each had at least 2 lactations. When unequal heritabilities were included for first, second, and third lactations, second and third lactations added little to first lactations in accurately calculating a cow's breeding value from records of her dam. When heritability was equal for all lactations (apparent lactation differences in heritability

caused by maternal effects) second and third records of the dam increased the correlation between genotype and calculated breeding value of the cow.

Butcher and Freeman (1968) analyzed 60,000 D.H.I.A. Holstein records and reported higher heritabilities for first lactations than for second lactations. They also found correlations between consecutive records increased as cows aged, but correlations between nonconsecutive records decreased as cows aged. They concluded weighting lactations from different parities separately would increase accuracy of selection.

Molinuevo and Lush (1964) used daughter on dam regression and also found first lactation Holstein production more heritable than second or third lactations. They suggested first lactation records would be sufficient for proving dairy sires, but second and third records could be useful in estimating breeding values of cows. Their paper contains a list of publications evaluating heritabilities of different lactations. Most of those do not differ in conclusion.

Butcher (1973) evaluated the use of first lactation paternal half sisters, dam, and maternal grandsire's first lactation daughters to estimate a bull's progeny test. Records were standardized 305-2X-ME deviates from herdmates. Using 340 Holstein bulls entering A.I. at less than 37 months of age, he calculated estimated breeding values for each bull based on various combinations of the three relatives. He computed correlations between these pedigree estimated breeding values and the bulls' progeny test values (SON). Breeding value computed on the first record of the dam was more highly correlated with SON (.21) than other dam records or average of dam records (.15). Measuring breeding value on maternal grandsire and various combinations of records of the dam

increased the correlations with SON over using dam alone. However, using dam average and maternal grandsire (.22) was less accurate than using only dam first lactation and maternal grandsire (.27). Highest correlations were .43 between SON and sire breeding value and .47 between SON and pedigree index of three dam lactations, sire, and maternal grandsire. Expected correlations were .38 and .45.

Butcher (1973) also computed linear regression coefficients to apply to combinations of breeding value estimates of sire, dam, and maternal grandsire to predict SON. Multiple correlation coefficients were higher when dam first lactation was the only dam information used. The highest multiple correlation was .482 for an equation including sire, dam first lactation, and maternal grandsire. Disregarding the dam reduced the multiple correlation to .457. Butcher summarized from the literature estimates of correlations between SON and breeding values on sire, dam average, and dam first lactation.

Separating bulls by pedigree index into quartiles he found the average predicted difference (Plowman and McDaniel, 1968) for each quartile declined as the rank of the quartile declined. However, the PD average in each group was substantially lower than the average pedigree index for each group. He suggested most of this difference was due to genetic trend.

Butcher also noted an increase in mean PD of sires as number of sons in A.I. increased. However, mean breeding value of dams was unrelated to number of sons in A.I.

In a study of pedigree selection of Holstein bulls M. G. Freeman (1970) also found pedigree breeding values overestimated future progeny

test breeding values, and early records on dams were most highly correlated with sons' progeny tests. However, regressions of son's progeny test on breeding values of various relatives differed from those derived by Butcher. These comparisons are in Table 1 and Table 2.

	<u> </u>			
Dam	measure	Butcher	Freeman	Vinson
lst	rec	.21	.23	
2nd	rec	.16	.21	
1 &	2	.16	.24	
1&	2&3	.21	.19	
Avg	. of all	. 22	.16	.11

Table 1. Correlations between son's progeny test and breeding value of dam.

	able	2.	Regression	of	son'	s	progeny	test	on	various	relative	ès.
--	------	----	------------	----	------	---	---------	------	----	---------	----------	-----

Relative	Butcher	Freeman	VanVleck ABS	VanVleck EAIC	Vinson
Sire's progeny test	.40	.22	.16	.31	• 20
Son's pedigree BV	.69	- 39	.56	.65	.17
Mat. grandsire prog. test	.24	09	.24	.07	

VanVleck and Carter (1972) compared Estimated Daughter Superiority (EDS) of Eastern Artificial Insemination Cooperative (EAIC) and American Breeders Service (ABS) bulls with the bulls' progeny tests (SC (sire comparison) and PD). EDS estimates one half the breeding value of a bull calculated from his sire, dam, and maternal grandsire. Correlations of each of a bull's relatives with his progeny test were lower than expected.

Dam's records had little importance in predicting son's progeny. Maternal grandsire was the relative most correlated with PD, but was least correlated with SC on various combinations of relatives. Some regressions are in Table 2. Regressing on the three relative groups which compose EDS accounted for less variation in PD or SC than regressing on EDS. All multiple correlations were less than expected. VanVleck concluded that although pedigree selection does not meet expectations it is an effective selection tool. He speculated EDS may predict PD or SC more accurately if heritability less than .25 were used in calculating EDS.

VanVleck (1969) studied the relative selection efficiency in retrospect of selected young sires. Data were 541 Holstein progeny from matings contracted by EAIC. Relatives of the young sire included sire, dam, maternal sisters, maternal grandsire, maternal grandam. The most efficient index in actual genetic gain versus theoretical genetic gain included sire, dam, and maternal grandsire. Differences in efficiency between pedigree combinations were small. VanVleck noted indexes have overemphasized all relatives except sire and maternal sisters. He suggested examining emphasis placed on milk fat percentage and type relative to milk yield.

Using data from seven bull stude Vinson (1971) reported regression coefficients between breeding values for milk yield of son and sire, son and dam, and son and midparent (sic) to be 35% to 45% of expected regressions. His correlation for dam and regressions for sons are in Tables 1 and 2. All were lower than most estimates in those tables. Midparent was equivalent to a pedigree estimate for the son involving paternal sisters and dam only. The breeding values were computed for males using daughter

production and for dams using dams' own production, all records deviated from herdmates. Vinson explained the loss of worth from expected to realized a consequence of genetic trend, female culling, differences in genetic merit of herdmates, and special treatment of potential dams of A.I. bulls. The study overemphasized fat production and type compared to milk yield in selecting dams of A.I. sires.

Walton (1970) investigated the relationship between PD's of 140 ABS Dairy Progeny Test bulls and their dams' production deviated from herdmates. He divided the dams into five groups of 28 in ascending order of deviation milk yield. Dams of the lowest and two highest groups produced sons whose PD's averaged the lowest of the five groups. He concluded more emphasis should be placed on sires than dams of bulls.

Attempting to account for variable genetic merit of mates of sires, Morillo and Legates (1970) and Bereskin and Freeman (1965) reported no accuracy gained by including records of dams of progeny in computing sire breeding values based on progeny tests.

Thomson and Freeman (1970) constructed pedigree indexes with and without environmental correlations between ancestors for 176 A.I. sires. Data were 158,000 records of 62,000 cows in 450 herds from 38 states. The correlations between index and sire genotype were higher for deviated records than mature equivalent (ME) records, and higher for indexes with no environmental correlations than indexes with environmental correlations. They concluded deviating records removed herd-year-season effects. Lower correlations using ME records were caused by differences between heritabilities and repeatabilities. Correlations between pedigree breeding values and FD's were higher for indexes using deviated records than

ME records and were unaffected by environmental correlations when deviated records were used. Indexes using only the maternal side of the pedigree were more highly correlated with PD's than indexes using all information, while indexes using only paternal information were the least correlated with PD's.

Before constructing a selection index it must be decided which measurements are worth including. Lush (1947) has an excellent discussion of interpreting biometrical relationships to combine information from various relatives in an index to make genetic gain from selection on the index.

Skjervold and Odegard (1959) and Young (1961) have presented selection index formulae to calculate weights to be applied to combinations of relatives to estimate breeding values of animals.

Miller (1968) discussed the development of selection indexes and contributions of various researchers to that development. He also discussed the use of an index and described the procedures involved in the United States Department of Agriculture (U.S.D.A.) cow index. Once each year U.S.D.A. publishes a list of the "best" two percent of registered cows enrolled in the D.H.I.A. program based on milk and fat indexes which include information on the cows and their paternal half sisters. Miller notes this list is published so the cows listed may be considered as bull dams.

EATA is a score given a dairy cow which estimates half her genetic worth relative to other EATA's in her population. It may be developed in the following manner.

H is the aggregate genotype of the cow such that H = .5G where G is the additive genotype of the cow for milk production. EATA is similar to an index value I where:

$$I = \sum_{i=1}^{5} b_i (X_i - \mu_i)$$

 X_i is a random variable with mean μ_i . (μ_1 is the average of all X_1 's) X_1 = mean deviation yield of the cow.

= mean milk production of the cow expressed as 305-2X-ME lactation
 yield minus a regressed herdmate average which is defined later.
X₂ = mean deviation yield of the cow's dam.

 X_2 = mean deviation yield of the cow's paternal half sisters.

 X_4 = mean deviation yield of the cow's daughters.

 X_5 = mean deviation yield of the cow's maternal half sisters.

Regressed herdmate average = $BSA + \frac{N-1}{N+1}$ (HA-BSA)

BSA = Breed-season average

- HA = Herd-year-season average excluding the record being deviated. Two seasons per year: 1) May-September. 2) October-April. See Bereskin (1963) for explanation of these seasons.
 - N = Number of herdmates. If N is less than sixteen an adjacent herd-year-season is also included in HA.

This formula gives HA more weight when there are more herdmates. A detailed description of EATA procedures was outlined by Eastwood (1968).

 b_i 's are chosen simultaneously such that R_{IH} is maximum. Multiplying I by a constant or adding a constant to I will not change R_{IH} . Adding a constant to X_i will not change b_i , and selection on X_i alone will not change b_i . The b_i 's are then calculated by maximizing R_{IH} as follows.

$$R_{IH} = \frac{Cov(I,H)}{(V(I) V(H))^{\frac{1}{2}}}$$
(1)

$$\ln R_{\rm IH} = \ln Cov(I,H) - .51nV(I) - .51nV(H)$$
(2)

$$= \ln(.5\Sigma_{i} x_{i}^{\sigma}) - .5\ln(\Sigma_{ij} b_{j}^{\sigma} x_{i}^{\sigma}) - .5\ln(.25\sigma_{G}^{2})$$
(3)

For the kth equation:

$$\sum_{i=1}^{\Sigma b} \sigma_{X_{k}X_{i}} = \cdot 5\sigma_{X_{k}G} \left(\frac{\sum_{j=1}^{\Sigma b} b \sigma_{j}X_{j}X_{j}}{\cdot 5\sum_{i=1}^{\Sigma b} \sigma_{X_{i}G}} \right)$$
(5)

$$\hat{L}_{i} \hat{\sigma}_{x_{k}x_{i}} = .5 \sigma_{x_{k}} G \left(\frac{\sigma_{\tau}^{2}}{\sigma_{IH}} \right)$$
(6)

If E(I-H)² is minimum, then $\frac{\sigma_{I}^{2}}{\sigma_{IH}} = 1$ and the b_i's from minimizing E(I-H)²

are the same as maximizing R_{TH} .

$$\sum_{i=1}^{\infty} \sigma_{X_{k}X_{i}} = .5\sigma_{X_{k}G}$$
(7)

The b's for the index are solutions of the equations above in matrix form below.

$$\begin{bmatrix} \sigma_{x_{1}}^{2} & \sigma_{x_{1}x_{2}} & \sigma_{x_{1}x_{3}} & \sigma_{x_{1}x_{4}} & \sigma_{x_{1}x_{5}} \\ \sigma_{x_{2}x_{1}} & \sigma_{x_{2}}^{2} & \sigma_{x_{2}x_{3}} & \sigma_{x_{2}x_{4}} & \sigma_{x_{2}x_{5}} \\ \sigma_{x_{3}x_{1}} & \sigma_{x_{3}x_{2}} & \sigma_{x_{3}}^{2} & \sigma_{x_{3}x_{4}} & \sigma_{x_{3}x_{5}} \\ \sigma_{x_{4}x_{1}} & \sigma_{x_{4}x_{2}} & \sigma_{x_{4}x_{3}} & \sigma_{x_{4}}^{2} & \sigma_{x_{4}x_{5}} \\ \sigma_{x_{5}x_{1}} & \sigma_{x_{5}x_{2}} & \sigma_{x_{5}x_{3}} & \sigma_{x_{5}x_{4}} & \sigma_{x_{5}}^{2} \\ & A & B & = D \\ B &= A^{-1}D \end{bmatrix}$$

Matrix A contains phenotypic variances and covariances between relatives and D contains covariances between relatives and half the genotype of the cow being indexed. $\sigma_{X_1}^2$ for instance is the variance of means of records per cow which have been deviated from herdmates. Whether X_1 has been deviated from the mean of all X_1 's (μ_1) is not important since $\sigma_{X_1}^2 = \sigma_{(X_1^- \mu_1^-)}^2$. Elements of D may be referred to as Right Hand Sides.

Matrices A and D are constructed from biometrical relationships between the five relative classes and parameters from the population, h^2 = heritability of deviated records, and r = repeatability of deviated records. Lactation j of individual i is represented by P_{ij} . P_{ij} is a

deviation from herdmates.

$$P_{ij} = g_i + c_i + e_{ij}$$

Here g_i is the genotype of cow i, c_i is the permanent environmental effect on cow i, and e_{ij} is all else contributing to record P_{ij} . The necessary variances and covariances are constructed from this model. The following relationships are declared to enable us to estimate variances and covariances and solve the equations. Some of these may not be biologically accurate.

1. Genetic differences are additive.

2. c, is uncorrelated with g's, e's, and other c's.

3. e_{ij} is uncorrelated with g's and other e's. Expanding all variances and covariances and dividing each by σ_g^2 yields the EATA equations in Figure 1. m_1 and m_2 are numbers of records in the averages of the cow being indexed and her dam. m_3 , m_4 , and m_5 are the average numbers of records for each of n_3 paternal half sisters of the cow, n_4 daughters of the cow, and n_5 maternal half sisters of the cow. For EATA's in these data $h^2 = .25$, r = .5. Minimum values for diagonal elements $\frac{\sigma_x}{\sigma_z}^2$, $\frac{\sigma_x}{\sigma_z}^2$, $\frac{\sigma_x}{\sigma_z}^2$, $\frac{\sigma_x}{\sigma_z}^2$, $\frac{\sigma_x}{\sigma_z}^2$ are respectively 2, 2, .25, .25, .25.

Average numbers of records, m_3 , m_4 , m_5 , are used to compute $\sigma_{x_3}^2$, $\sigma_{x_4}^2$, and $\sigma_{x_5}^2$. Including specific numbers of records for each individual in the average will give variances at least as large as those calculated for average numbers of records. Since the covariances between mean production of related individuals are unaffected by numbers of records in those means, the amount by which $\sigma_{x_i}^2$ is underestimated is a function of variances of means. Where S_i is the number of records of the ith cow in a group of n relatives, the amount by which $\sigma_{x_i}^2$ is underestimated is:

$$\left(\frac{1}{\frac{1}{n^{2}}} \sum_{i} \frac{1+(S_{i}-1)r}{S_{i}} - \frac{1+(\sum_{i} S_{i}/n-1)r}{\sum_{i} S_{i}}\right) \sigma_{p}^{2}$$

The size of this value depends on the number of relatives in the group and the distribution of numbers of records per individual. For the following three paternal sister cases I have calculated the percentage of the variance now used which needs to be added to increase it to the variance calculated without averaging the numbers of records.

Cas	<u>e 1</u>	Case	<u>e 2</u>	Case 3			
No. PHS	Recs/PHS	No. PHS	Recs/PHS	No. PHS	Recs/PHS		
500	1	2	1	1	1		
200	3	1	5	1	11		
200	6	3 avg.	2.3	2 avg	. 6		
100	10						
1000 avg	. 3.3						
.2	7.	18	Z		31%		

It appears the variances involving average numbers of records are too small when cows are few and records per cow vary widely. In those instances paternal sisters, daughters, or maternal sisters would receive more weight in the index than they deserved.

For each cow to be indexed the appropriate m's and n's are inserted in the five equations and the equations are solved for the b's. Those b's are then used in the index equation along with the mean deviation production of the cow's relatives to calculate an EATA for the cow.

EATA =
$$b_1 x_1 + b_2 x_2 + b_3 x_3 + b_4 x_4 + b_5 x_5$$

Note this equation is EATA = $\Sigma b_i X_i$ where the original index equation was $I = \Sigma b_i (X_i - \mu_i)$. In the EATA equation the means are ignored for ease of computation. Note μ_i will be close to zero since it is the mean of deviated records. This situation and its consequences will be discussed later. If information on one or more relative groups is not available, the b's are computed from an abbreviated set of equations where rows and columns pertaining to the missing relatives have been deleted from the variance-covariance matrix A, and the corresponding b's and covariances have been deleted from the B and D vectors. Those b's and associated X's are also deleted from the EATA equation.

An example illustrating computation of an EATA from Iowa D.H.I.A. data is in Table 3. EATA represents half the breeding value of the cow, and although its major use is ranking cows for selection, it can be added to the Predicted Difference of a sire to estimate the breeding value of an offspring from that mating.

R_{IH}, the correlation between the index and the aggregate genotype, can be expressed computationally as:

$$R_{IH} = \left(\frac{\sum b_i RHS_i V(G)}{V(.5G)}\right)^{\frac{1}{2}}$$



Figure 1. Equations for computing weights for EATA.

i	Relative	No. of	Animals	No. of Records/ Animal	X = Avg. Yield(kg)/ i Animal	b _i
1	Cow			7	1515	.15
2	Dam			5	435	.06
3	PHS	435		3.7	319	.31
4	Daughters	2		2.0	736	.07
5	MHS	1		5	-550	.02
	EATA R _{IH}	$= b_1 X_1$ $= .15(1)$ $= \left(\frac{b_1 RH}{.15(1)} \right)$ $= \left(\frac{.15(1)}{.15(1)} \right)$ $= (.59)$ $= .77$	$+ b_2 x_2 + .0$ 515) + .0 227 + .0 393 $35_1 + b_2 RH$.5) + .06 $()^{\frac{1}{2}}$	$b_{3}x_{3} + b_{4}x_{4} + b_{5}x_{5}$ 6(435) + .31(319) 26 + 99 $S_{2} + b_{3}RHS_{3} + b_{4}RH$.25 (.25) + .31(.125) .25	$\frac{5}{4} + .07(736) + .02(-550) + 52 - 11$ $\frac{45_4 + 5_5 RHS_5}{4}^{\frac{1}{2}} + .07(.25) + .02(.125)$) ¹ 2

.

ı

Table 3. Computation of an EATA.



 RHS_i is the element of the D matrix corresponding to b_i . An R_{IH} is computed in Table 3. R_{IH} will not change from one EATA to another providing the combination of relative records does not change. If the combination of records changes then R_{IH} may change. This refers to R_{IH} 's based on the three theoretical properties declared previously.

Possible inaccuracies in the use of EATA to rank cows may fall into four categories.

- 1) Development and use of EATA index equations.
- Calculation of parameters and relationships for the equations.
- 3) Measurement of milk yield deviated from herdmates.
- 4) Mathematical errors.

Categories two and three will be discussed later relative to results of this project. The first topic however pertains more closely to this section and will be discussed here.

When maximizing R_{IH} the ratio $\frac{\sigma_1^2}{\sigma_1}$ was substituted as 1 under the

restriction that $E(I-H)^2$ was minimum. What b's would be obtained from maximizing R_{IH} without restricting $E(I-H)^2$ to be minimum? Starting from the point of the restriction, equation k from (5) is:

For ease of computation let i = 1, 2. The equations for k=1 and k=2 then reduce to:

$$b_{1}(\sigma_{x_{1}}^{2}\sigma_{x_{2}}G^{-}\sigma_{x_{1}}x_{2}^{\sigma}x_{1}G^{-}+b_{2}(\sigma_{x_{1}}x_{2}^{-}\sigma_{x_{2}}G^{-}\sigma_{x_{2}}^{2}\sigma_{x_{1}}G^{-}=0$$

$$\frac{b_{1}}{b_{2}} = \frac{\sigma_{x_{2}}^{2}\sigma_{x_{1}}G^{-}\sigma_{x_{1}}x_{2}^{-}\sigma_{x_{2}}G^{-}\sigma_{x_{2}}G^{-}\sigma_{x_{2}}G^{-}\sigma_{x_{1}}G^{-}\sigma_{x_{2}}G^{-}\sigma_{x_{1}}G^{-}\sigma_{x_{2}}G^{-}\sigma_{x_{1}}G^{-$$

 R_{IH} is maximized when the ratio $\frac{b_1}{b_2}$ equals the expression above. There is

no unique solution for b's to maximize R_{IH} . Only the ratio of b's is unique for a specific combination of relatives and records. Restricting $\frac{\sigma_{I}^2}{\sigma_{IH}} = 1$, which implies E(I-H)², is minimum is one of an infinite number of

restrictions. Other reasonable restrictions will be discussed later.

A potential problem with EATA is that the means of the deviated records are not known. EATA is related below to I, the index defined previously.

> $I = \sum b_i (X_i - \mu_i)$ EATA = $\sum b_i X_i$ EATA = $I + \sum b_i \mu_i$

If $\Sigma b_{i} \mu_{i}$ is constant for all EATA's then differences between EATA's will be the same as differences between I's. If the μ 's are zero or if the μ 's are equal and the b's sum to zero then I = EATA. EATA's with identical b's differ the same as their corresponding I's differ. A problem develops when none of these relationships holds and the b's vary from one EATA to another. In this case $\Sigma b_{i} \mu_{i}$ changes from cow to cow because of changes in numbers of records and relatives. μ_{i} 's remain constant from cow to cow because they are population means. Consider the following two EATA's.

EATA₁ = $I_1 + \Sigma b_i \mu_i$ EATA₂ = $I_2 + \Sigma b_i^* \mu_i$ EATA₁ - EATA₂ = $I_1 - I_2 + \Sigma b_i \mu_i - \Sigma b_i^* \mu_i$

If $\Sigma b_{\mu} \neq \Sigma b_{\mu}^{\mu}$ then the difference between the cows' EATA's is not the same as the difference between their I's. They may change rank, a circumstance which could cause an incorrect selection decision.

To remedy the problem of unknown means and differing b's Henderson (1963) has suggested assuming all μ_i equal and forcing $\Sigma b_i = 0$. This causes $\Sigma b_i \mu_i = 0$ and E(EATA) = 0. The EATA is then an unbiased estimator of $\frac{1}{2}$ the cow's genic worth. He outlines two methods of accomplishing this, one using a Lagrange multiplier equation and one using a maximum liklihood estimate of μ from the X_i in the index. Both methods give the same solutions for the b's. Henderson comments that this method maximizes R_{TH} subject to $\Sigma b_i = 0$.

This is correct subject to $\Sigma b_i = 0$, but that alters the most important property of EATA, that R_{IH} be maximum. R_{IH} , when $\Sigma b_i = 0$, will be substantially lower than that attained with the usual b's. Three facts make this evident. First, when $\Sigma b_i = 0$ it is impossible to maintain the

constant ratios of b's required to maximize R_{IH} as shown earlier in this section. The ratios will not remain constant when some b's are positive and some negative, which occurs if any b is not zero and their sum is zero. Second, solutions to the EATA equations were unique without restrictions and would maximize R_{IH} . Restricting solutions to these equations without maintaining proportionality among the b's reduces R_{IH} . Third, augmenting the original EATA equations when they contain no dependencies changes the equations such that the biometrical relationships among the right hand sides are altered. RHS₁ becomes (RHS₁-a) where <u>a</u> is the Lagrange multiplier.

The following example will illustrate some consequences of restricting $\Sigma b_i = 0$. I will rank three cows by their values calculated from three different indexes.

1) Index to maximize R_{IH} where μ_i 's are known and equal. (Index)

2) EATA index where means are unknown. (EATA)

3) EATA index subject to $\Sigma b_i = 0$, means unknown but equal. (R. Index) Each of the three cows has the same information available, five records of the cow, four records of her dam, and 100 paternal sisters with two records each.

			Cow 1	<u>Cow 2</u>	<u>Cow 3</u>
x ₁	= A f	verage yield of the cow devia from herdmates.	ated 400	50	0
х ₂	= A V	verage yield of the dam de- viated from herdmates.	100	110	-20
х ₃	= / v	verage yield of the PHS de- viated from herdmates.	120	0	100

 μ_1 = population mean of X_1 = 10 μ_2 = population mean of X_2 = 10 μ_3 = population mean of X_3 = 10

Equations to calculate b's for Index and EATA:

$$\begin{bmatrix} 2.4 & .5 & .25 \\ .5 & 2.5 & 0 \\ .25 & 0 & .28 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} .5 \\ .25 \\ .125 \end{bmatrix} \qquad b_1 = .163$$

$$b_1 = .163$$

$$b_2 = .067$$

$$b_3 = .301$$
Cow 1 Index₁ = .163(400-10) + .067(100-10) + .301(120-10) = 103

Cow 2 Index₂ = .163(50-10) + .067(110-10) + .301(0-10) = 10 Cow 3 Index₃ = .163(0-10) + .067(-20-10) + .301(100-10) = **2**3

Equations to calculate b's for Restricted Index ($\Sigma b_i = 0$).

	5 /		5	25	T	$\Gamma_{\rm h}$]	L J		h	= 16	<u>م</u>	
	2.4	•	• •	•.6.7	1	11			51	10	,,	
	.5	•	2.5	0	1	^b 2	- 25		^b 2	= .03	L3	
	. 25	•	0	. 2 8	1	^b 3	125		^ь з	=18	32	
		-	1	1	0	a	0		a	= .13	34	
Cow	1	R.	Index ₁	= .169	(400)	+ .03	L3 (10	0)18	82	(120)	"	47
Cow	2	R.	Index ₂	= .169	(50)	+ .01	13 (11	0)18	82	(0)	=	10
Cow	3	R.	Index,	= .169	(0)	+ .03	13 (-2	0)18	82	(100)	=	-18

The Index using known means is the most accurate index. In this example all cows have equal information (identical b's) and $\mu_1 = \mu_2 = \mu_3$. Yet problems specific to the Restricted Index surface immediately. Although the expected value of the Restricted Index is zero and R_{IH} is maximized subject to $\Sigma b_1 = 0$, it gives negative weight to paternal sister information thereby reversing the rank of cows two and three, and reducing R_{IH} to less than 70% of its value in the Index. Because of the equal information the EATA's differ from the Indexes by a constant and maintain the .74 R_{TH}.

The unreasonableness of using $\Sigma b_i = 0$ to rank cows can be further illustrated. If there is more than one b, at least one has to be negative which means information on a relative positively genetically correlated with the cow is subtracted. This is not an impossible situation for all indexes in that a variable may receive negative weight if the phenotypic correlation between relatives is larger than the genotypic correlation (see Lush, 1947). Negative weights are impossible to maximize R_{IH} with EATA since genetic correlations are $1/h^2$ as large as phenotypic correlations when environmental correlations are not considered. If there is information on the cow only, there is only one b and it has to be zero. This creates a situation where every such cow regardless of production is worth zero and has an equal chance of being selected.

Further investigation into the numeric example makes it apparent the following three provisions are not compatible.

1. $\Sigma b_i = constant$.

2. Proportionality of b's maintained for maximum R_{TH} .

3. Right Hand Sides maintain correct biometrical relationships. This is evident with the following augmentation of the index equations used in the example. Solution b's of these equations maintain the same proportionality to each other as did the Index b's when the last element in the Right Hand Sides is a positive constant (h² here).

2.4	•2	.25	1	b ₁	.5]	b ₁ = .063
.5	2.5	0	.5	b ₂ =	.25	b ₂ = .026
. 25	0	.28	. 25	Ъ ₃	.125	b ₃ = .117
2.4	2.5	. 28	0	a	.25	a = .306

In this example I maintained the proportionality of the b's as stated, and I maintained the correct relationships between the Right Hand Sides by adding a, .5a, and .25a to each equation respectively. To satisfy those two properties I have a situation where $\sum b_i$ will not equal the same constant if I change the numbers of records. In the fourth equation I made the sum of b's times their respective diagonal elements equal a constant (.25). This could be done for every index with different numbers of records as long as the last equation were $\sum diagonal element_i$

x b_i = .25. Having this property for all indexes may have value, but it escapes me. This has not excluded the possibility that some augmentation

¹I term this Equal Opportunity Indexing.

will cause $\sum b_i$ = constant and b's proportional without the biometrical relationships of Right Hand Sides preserved.

The problems of comparing index values of different cows may be summarized briefly. When the μ 's are included in the index the expected value of the index, E(I), equals zero, and the variance of the index, V(I), varies from one combination of relatives and records to another as the b's and X's change. When the means are not included in the index E(I) varies as the b's change from cow to cow, and V(I) continues to vary as it did when the means were included. R_{IH} is maximum with or without the means. A solution to the problem of correctly indexing cows for ranking and selection lies in attaining the following properties, listed in order of priority highest to lowest.

 Maximize R_{IH}, which involves maintaining the appropriate ratios among the b_i's for each combination of relatives and records.

2. Make E(I) constant.

3. Make V(I) constant.

These would maximize the correlation between index and cow's genotype, measure every cow from the same base, and standardize the spread in indexes regardless of the amounts of information involved. Developing these three properties in one index is a thesis topic itself. However, since the only changes to an index which will not affect the ratios of the b_i's are to multiply the index by a constant or add a constant, they are good prospects for deriving an "ultimate" index.

Much discussion has been devoted to the potential problems of not knowing the means of the X_i 's when computing indexes. It seems reason-

able to calculate consequences of this situation with respect to EATA and selection decisions. I wrote previously the difference between two EATA's differs from the difference between their correct indexes by the quantity now d.

$$\mathbf{d} = \Sigma \mathbf{b}_{\mathbf{i}} \boldsymbol{\mu}_{\mathbf{i}} - \Sigma \mathbf{b}_{\mathbf{i}}^{\mathbf{i}} \boldsymbol{\mu}_{\mathbf{i}}$$

The question becomes, how large is this quantity. Will it change the rank of two cows in such a circumstance that the genetically inferior cow will be saved and the superior cow discarded? Or will a bull calf be discarded when he should be saved?

If all cows included in an HA are indexed with EATA, all μ 's will equal zero, d will be zero, and differences between cows will be accurately reflected in their EATA's. If not all cows are indexed, the population on which selection will be practiced differs from the original population. This occurs when cows who have contributed records to herdmate averages are dead or off test and not indexed. The population mean of the smaller population may not equal the original population mean (zero) if selection has been practiced. Since only a small fraction, less than 15%, of the cows not indexed would have been culled for low production the new mean deviation yield would be slightly greater than zero. For instance the mean yield average deviated from herdmates for daughters of live registered Holsteins in 1973 in Iowa, Kansas, and Missouri was fifteen kg. Those daughters have been highly selected on dam longevity for whatever that involves, and I consider their production an upper limit of the value of $\boldsymbol{\mu}_{i}$ for these EATA's. If each $\boldsymbol{\mu}_{i}$ is between zero and fifteen kg then d will be less than

$$d < 15(\Sigma_{1} - \Sigma_{1})$$

For an indication of a likely maximum size of d we need estimate a maximum value for $\Sigma b_i - \Sigma b_i^{!}$. The information on two cows in Table 4 will nearly maximize this value. Using .70 from the table, a reasonable maximum value of d is:

$$= 10.5 \text{ kg}$$

Table 4. Example of large $(\Sigma b_i - \Sigma b_i)$.

		Cow 1			Cow 2				
Relative	No.	Recs.	bi	No.	Recs.	^b i			
1 Cow		15	.10		1	.12			
2 Dam		15	.02						
3 PHS	50,000	5	. 20						
4 Daus.	15	5	.40						
5 MHS	15	5	.10			<u></u>			
	$\Sigma b_i =$.82		Σb _i =	.12			
$\Sigma b_i - \Sigma b_i = .8212$									

Considering milk yield is recorded to the nearest ten pounds (4.5 kg), misevaluating by \pm 10.5 kg at most the true difference between two cows is a minor error. Under these circumstances few mistakes will be made by culling superior cows. If the mean is closer to zero than fifteen and if cows being compared have more nearly equal information than the two
in Table 4, the selection errors made by ignoring the μ_i 's in EATA will have little consequence on genetic progress. The expense of changing the current EATA program to account for the μ_i 's almost certainly would exceed the profit gained by more accurate cow rankings.

In recent years Henderson has approached sire evaluation through linear equations which yield estimates with minimum errors of prediction among unbiased linear estimators. Applying that procedure to cow evaluation is now of interest to many researchers.

HISTORY OF EATA

Since EATA is a selection index procedure its history traces to initial efforts to apply indexes to biological species.

Smith (1936) developed Fisher's concept of discriminant functions for use in constructing selection indexes for plants. Application of selection indexes to animal populations was described by Hazel (1943) in his paper "The genetic basis for constructing selection indexes."

Lush developed in the 1930's and applied routinely in his breeding research an index of information from relatives to select cows and bulls on yield and type.

In 1953 L. McGilliard roughed out selection index weights to rank dairy cows on their genetic worth. The five relative groups currently used in EATA were included in the index and the measure of milk yield was an average for each female. Weights for each relative were independent of other relatives. The first ranking for institution herds in Iowa in 1953 was a dual ranking with all differences between herds genetic and, the other extreme, all differences environmental.

The 1955 rankings were based on new weights developed by McGilliard from relationships from a variety of analyses. A type score was also included. Weights for relatives were not independent. The measure of production was fat-corrected 3.5 average deviations from yearly averages. They were constructed as if all differences between herds were environmental because analyses indicated that was the case for the institution herds.

After 1956 Freeman continued rankings for the institution herds. In the sixties many people worked on the index to incorporate paternal sister weights correlated with other relatives, regressed herd-year-season averages, and other changes. In the winter of 1967-68 Eastwood provided the first EATA rankings to dairymen in the Postville, Iowa area. Then in November 1968 rankings were provided for all herds on test in the eight state region. The term "EATA" was chosen for the listing because "ETA" sounded too similar to "EPA" which was another ranking already in existence. The word "Average" was included to emphasize the concept of this being the cow's average contribution transmitted.

Between 1968 and 1970 EATA rankings were computed at yearly intervals except for two years at six month intervals. When funds became scarce, 1.5 years elapsed after the 1970 ranking before the next came out in May 1972. Yearly rankings were then distributed in May until 1974. EATA's will now be computed annually beginning September 1974. These EATA's will include changes to account for average estimated genetic worth of herdmates inadvertently being removed from the deviations with the regressed adjusted herd-year-season averages.

Individual lactation records used in this study were Dairy Herd Improvement Association "1095" records restricted to:

DATA

Holstein breed Official test Iowa, Missouri, Kansas Active herds January 1967 Calving dates since January 1, 1967 Lactations completed by January 1973.

So pedigree information would be available for every cow, the 471, 913 records satisfying the above requirements were restricted to records of registered cows. There were 121,299 of those records.

Cows with EATA records computed by D.H.I.A. were restricted to cows with EATA's computed May 1, 1972 which meant they must have been alive after November 1, 1970. EATA's on older dead cows were not available. Cows also had to be sire identified or dam identified or have at least one daughter. There were 94,487 of these EATA's of which 36,475 represented registered cows.

Each individual lactation of a cow was paired with the EATA information of the cow's dam. There were 20,414 lactations in 855 herds where the dam of the cow had an EATA. This later became 20,377 lactations as 37 were lost. Information on the sire of the cow was also included when available from a January 1973 list of 2392 sire PD's.

Since the goal of this study was to evaluate how predictive of daughter production a dam's EATA is, it was important the EATA did not include

records of the daughter. Including the predicted daughter record in dam EATA would nearly double the expected covariance between them. Only 423 records in progress less than 45 days as of May 1, 1972 were available and had not been included in the dam EATA's. Therefore all other records of each daughter to be predicted were removed from their dam's EATA, and the EATA and daughter information in it were adjusted to what they would have been had the daughter to be predicted been without records.

A record in this study then consisted of an individual lactation and information about a cow, EATA information on her dam with that specific cow not included, and PD information on her sire. Milk weights were recorded in tens of pounds. All pounds were converted to kilograms before analysis. Information available for a cow and her dam was:

Cow

Cow Registration Number Sire Registration Number Dam Registration Number Birth Date Calving Date Days in Milk (first 305) Milk Fat Termination Code State, County, Herd Lactation Number Cow Index Number ME Mill ME Fat Average Herdmate ME Milk Average Herdmate ME Fat Number of Herdmates

Dam

Active in Herd Sire Registration Number Dam Registration Number Cow Records (Number) Cow Average Milk (ME deviated) Cow Average Fat Dam Records (Number) Dam Average Milk Number of Paternal Half Sisters Average records per PHS (Number) PHS Average Milk PHS Average Fat Number of Daughters Average records per Daughter (Number) Daughter Average Milk Daughter Average Fat Number of Maternal Half Sisters Average Records per MHS (Number) MHS Average Milk MHS Average Fat EATA Milk EATA Fat

METHODS

The following equation was solved to minimize the contribution of error to differences between daughters.

$$Y_{ijk} = \mu + S_i + b(EATA_j - \overline{EATA}) + e_{ijk}$$

 Y_{ijk} is lactation milk production of the kth daughter of the ith sire and jth dam. μ is the mean of Y_{ijk} 's. S_i is the effect of the ith sire. b is the regression of daughter production on dam EATA calculated within sires. The μ + S_i were absorbed (more precisely described below) to remove effects of daughters' sires from differences between daughters. Differences which remained were due to dams and residual variation. EATA_j is the EATA of the jth dam not including records of the kth daughter. EATA is the mean of all dam EATA's. e_{ijk} is error peculiar to Y_{ijk} . Y_{ijk} was expressed as kilograms of 305 day milk production, or 305 day production adjusted to what the cow would have produced in 305 days under identical conditions had she been a mature cow, milked twice daily, during an average season (Mature Equivalent or ME). Or Y_{ijk} was expressed as ME production deviated from herdmates' ME average, which removes effects common to a cow and her herdmates. Common effects may be environmental and genetic.

Quadratic and cubic equations were also solved. They were: Quadratic: $Y_{ijk} = \mu + S_i + b_L(EATA_j - EATA) + b_Q(EATA_j^2 - EATA^2) + e_{ijk}$ Qubic: $Y_{ijk} = \mu + S_i + b_L(EATA_j - EATA) + b_Q(EATA_j^2 - EATA^2) + b_Q(EATA_j^2 - EATA^2) + b_Q(EATA_j^3 - EATA^3) + e_{ijk}$

 b_L , b_O , b_C are linear, quadratic, and cubic partial regressions of daughter

production on dam EATA, within sire of daughter.

After the regressions of daughter production on dam EATA, I regressed daughter production simultaneously on the five groups of relatives in a dam's EATA. That equation was:

 $Y_{ijk} = \mu + S_i + b_1 Cow_j + b_2 Dam_j + b_3 PHS_j + b_4 Daus_j + b_5 MHS_j + e_{ijk}$ Y_{ijk} , μ , S_i , and e_{ijk} have been defined. All regressions were computed within sire. Each b is the partial regression of daughter milk production on production of the relatives designated following each b. Cow_j represents average of records of the dam of the daughter. Dam_j represents average of records of the dam of cow_j . PHS_j represents the average of records of paternal half sisters of cow_j . Daus_j represents average of records of daughters of cow_j , excluding the daughter who produced record Y_{ijk} MHS_j represents the average of records of maternal half sisters of cow_j . Records of the independent variables were expressed as ME records deviated from herdmates.

The μ + S_i were solved in terms of other variables in the equations and substituted into the original equations. These were solved for the b's without numerically evaluating the μ + S_i. Equations involving EATA will be termed EATA equations. Those involving components of EATA will be termed components equations.

Expected Regressions and correlations

X represents a daughter record with sire variation removed and may be expressed as:

$$X = .5G_{D} + (.5)^{\frac{1}{2}}G_{M} + E$$

 G_D is the genetic worth of the dam. G_M is the genetic contribution to the daughter from Mendelian sampling of genes of the sire and dam. G_M will average zero, but will contribute half the genetic variation of daughter records unadjusted for sire. E is non-genetic contributions to the daughter ter record. EATA of the dam may be expressed as:

EATA =
$$.5 G_D$$

The variances and covariance of Y and EATA are the following. $V(G_D) = V(G_M) = \sigma_G^2$. Cov (G,E) = 0. $V(X) = .75 \sigma_G^2 + \sigma_E^2$. $V(EATA) = .25 \sigma_G^2$. Cov $(X,EATA) = .25 \sigma_G^2$. The correlation between daughter record (intrasire) and her dam's EATA is expected to be ρ . The intrasire regression of daughter record on dam's EATA is expected to be β . Both are developed as follows:

$$\rho = \frac{\text{Cov}(X, \text{EATA})}{(V(X) \quad V(\text{EATA}))^{\frac{1}{2}}}$$

$$= \frac{.25 \sigma_{G}^{2}}{(.75 \sigma_{G}^{2} + \sigma_{E}^{2}) (.25 \sigma_{G}^{2})^{\frac{1}{2}}}$$

$$= \frac{.25 h^{2}}{(.75 h^{2} + 1 - h^{2}) (.25 h^{2})^{\frac{1}{2}}}$$

$$= \frac{.0625}{((.9375) (.0625))^{\frac{1}{2}}} \quad (\text{if } h_{2} = .25)$$

$$= \frac{.0625}{.2421}$$

$$\beta = \frac{Cov(X, EATA)}{V(EATA)}$$
$$= \frac{.25\sigma_{G}^{2}}{.25\sigma_{G}^{2}}$$

 ρ may also be written as $\left(\frac{.25 \text{ h}^2}{1 - .25 \text{ h}^2}\right)^{\frac{1}{2}}$, where the variation in daughter records accounted for by the dam is $.25 \sigma_G^2$ and the variation in daughter records where sire variation has been removed is $\sigma_x^2 - .25\sigma_G^2$.

The expected values of this correlation and regression may also be expressed in terms of components of the dam's EATA. This will account for the combination of relative information in an average EATA. This process is described below.

EATA = A(COW) + B(DAM) + C(PHS) + D(DAUS) + E(MHS)A, B, C, D, E are weights derived from the theoretical EATA equations using the numbers of relatives and records in an average EATA. Other symbols were defined previously.

Cov (X, EATA) =
$$.5 \ A\sigma_{G}^{2} + .25 \ B\sigma_{G}^{2} + .125 \ C\sigma_{G}^{2} + .25 \ D\sigma_{G}^{2} + .125 \ E\sigma_{G}^{2} + .125 \ E\sigma_{G$$

 $V(EATA) = A^2 V(COW) + B^2 V(DAM) + C^2 V(PHS) + D^2 V(DAUS) + E^2 V(MHS)$ + 2AB Cov(COW, DAM) + 2AC Cov(COW, PHS) + 2AD Cov(COW, DAUS) + 2AE Cov(COW, MHS) + 2BC Cov(DAM, PHS) + 2BD Cov (DAM, DAUS) + 2BE Cov(DAM, MHS) + 2CD Cov(PHS, DAUS) + 2CE Cov(PHS, MHS) + 2DE Cov(DAUS, MHS) A more detailed expansion of these variances and covariances is in Figure 1. Substituting heritability of .25, repeatability of .5, and the 38

average numbers of relatives and records in EATA's of dams of first lactation daughters as in Table 5, the expected correlation and regression become:

for EATA.	,,,,,	·····
A = .15		$m_1 = 5.1$
B = .06		$m_2 = 3.7$
C = .33	$n_3 = 1044$	$m_3 = 2.5$
D = .03	n ₄ = .9	$m_4 = 1.2$
E = .02	$n_5 = 1.4$	$m_5 = 2.2$

Table 5. Average numbers of relatives, records, and associated weights

ρ=	Cov (X, EATA)
	$(V(X) V(EATA))^{\frac{1}{2}}$
=	.0353
	((.9375)(.0349)) ¹ 2
=	<u>.0353</u> .1808
=	.195
β =	$\frac{Cov(\chi, EATA)}{V(EATA)}$
=	.0353 .0348
=	1.01

The expected variance of EATA based on average information is only half the variance of .5 G_D which EATA estimates. β of one indicates a unit difference between dam EATA's will show up as a unit difference between their daughters within sire. This is reasonable since EATA represents half the dam's genetic value which she passes to her offspring.

 ρ^2 is the expected fraction of variation in daughter records (intrasire) accounted for by variation in their dams' EATA's. It is the fraction of differences between daughter records which disappears when dam EATA is held constant. ρ^2 is expected to .067 for unlimited information in EATA and .038 for the information in these data.

RESULTS AND DISCUSSION

A summary of animal numbers and lactation yield for first, second, and average of all daughter lactations is in Table 6. Daughters averaged 1.8 lactations each. The increase of daughter production from first to second lactation was probably a consequence of culling. Standard deviations are in parentheses. Standardizing and deviating records from herdmates reduced the variation between cows. Had all cows been included regardless of availability of dam EATA, deviation yield would have averaged zero. Both dam EATA and deviation yield averaging above zero indicates cows missing dam EATA's would have averaged less production than their herdmates. EATA's without daughters also would have averaged below zero.

Correlations between dam EATA and daughter milk yields are in Table 7. The expected correlation of daughter first lactation deviated production with dam's EATA is .20 and the .18 obtained does not differ significantly (P < .05) from it. The expected correlation would be lower for ME yield and lactation yield as differences between daughters increase with these measures. The correlation higher than expected was between dam EATA and daughter ME production. This resulted from a covariance 29% larger than the covariance between dam EATA and deviated daughter production. Since the only difference between those measures of daughter production was subtracting the herdmate ME average, the covariances between EATA and daughter production were expected to be equal. The inequality and size of covariances indicate a positive correlation between dam EATA and herdmate average of daughter, which arises when managers of some herds cull more

	Daughter Lactations						
	lst	2nd	Average of All				
Number of Daughters	10349	5415	11106				
Number of Sires	1 95 5	139 8	2159				
Dam EATA (kg)	34 (192)	39 (191)	34 (192)				
Dau. Lactation Yield (kg)	5028 (1_14)	5699 (1626)	5179 (1583)				
Dau. ME Yield (kg)	6619 (1372)	6758 (1523)	6624 (1344)				
Dau. Dev. Yield (kg)	7 (1240)	176 (1331)	15 (1206)				

Table 6. Numbers of animals and yield means.^a

^aStandard Deviations in parentheses

Table 7. Correlations between dam EATA and daughter yield.

	Daughter Lactations				
	lst	2nd	Average of All	Expected	
Lactation Yield	.16	.19	.16		
ME Yield	. 21	.21	. 2 1		
Dev. Yield	.18	.19	.18	. 20	

intensely or use a higher proportion of genetically superior sires than other herds. Higher correlations for daughter second lactations than for daughter first lactations indicate daughters surviving to produce a second record had more in common with their dams' EATA's than did average daughters, especially since second lactations were more variable than first lactations. Higher correlations were caused by larger covariances.

Partial regression coefficients from regressing daughter production on dam EATA for linear, quadratic, and cubic equations are in Table 8. R^2 is the fraction of variation in daughter records accounted for by the powers of EATA in the model. All linear regressions except one differed significantly (P<.05) from the expected 1.01. This means the probability of a regression that large when it does not differ from expected is less than .05. The exception was the first lactation linear coefficient in the quadratic model.

With all models the linear coefficients were largest for ME daughter records and lowest for deviated records. In a regression, variation in the dependent variable directly influences only the covariance. Therefore, differences in linear coefficients can be attributed to the covariance between daughter and dam EATA increasing when daughter records were ageseason adjusted and decreasing substantially when these ME records were deviated from herdmates. This implies high dam EATA's were associated with high daughter herdmate ME records, as previously discussed. Second lactation linear regressions were larger than first lactation due to selection also previously discussed.

Nearly all quadratic partial regressions for actual and ME records were positive and differed significantly from zero. Deviating them from

Measure of Daughter Yield	Linear ^a	Quadratic	Cubic	R ²
lst Lactation	1.21 (.08)*			.024
lst ME	1.48 (.08)*		·	.043
lst Dev.	1.15 (.07)*		·	.031
lst Lactation	1.15 (.09)	.0006*		.025
lst ME	1.42 (.08)*	.0005*		.043
lst Dev.	1.16 (.07)*	0001		.031
2nd Lactation	1.53 (.14)*	.0006		.036
2nd ME	1.61 (.13)*	.0008*		.047
2nd Dev.	1.30 (.12)*	0001		•034
Avg. Lactation	1.24 (.09)*	.0006*		.026
Avg. ME	1.44 (.08)*	.0005*		.046
Avg. Dev.	1.16 (.07)*	0002		.033
lst Lactation	1.29 (.12)*	.0009*	12 x 10 ⁻⁵	.025
lst ME	1.61 (.11)*	•0009*	16 x 10 ^{-5*}	.044
lst Dev.	1.26 (.10)*	.0001	09×10^{-5}	.032
Expected 1st Dev.	1.01	0	0	.038

Table 8. Regressions of daughter yield on dam EATA.

*Differs significantly from expected (P<.05). Probability of a regression this large when there is no difference from expected is less than .05.

^aStandard errors in parentheses.

herdmates reduced them to zero.

Qubic partial regressions were negative but differed from zero only for ME records.

 R^2 's were largest for ME records and smallest for actual records. Actual records were the most variable. R^2 's were slightly lower than expected for deviated first lactation production, and were of equal size regardless of model, linear, quadratic, or cubic. Statistically the graph of daughter first lactation deviated production versus dam EATA was linear with a slope steeper than expected.

Daughters of dams with high EATA's produced deviated records larger than daughters of dams with lower EATA's. A one unit difference in dam EATA's resulted in 1.15 units difference in daughter deviations. Linear, quadratic, and cubic graphs of this are Figures 2, 3, and 4. Figure 5 is a cubic graph of first lactation daughter ME production.

I was concerned dam EATA might be additionally useful to explain daughter variation if fluctuations in daughter production specific to certain magnitudes of EATA were being masked by the bulk of observations or the symmetry of a cubic equation. There was special interest in daughters of high EATA dams because they might become dams of A.I. sires. If males they would likely be A.I. sires. To investigate these concerns I divided all first lactation daughter records into eight groups, ordered by size of dam EATA, each group containing a one standard deviation (192 kg) range of EATA's. I then related daughter production to dam EATA within each EATA segment.

Table 9 contains numbers of daughters and sires and lactation milk yield means for dam EATA's and daughters within each EATA segment.



Figure 2. Linear regression of daughter first lactation deviation on dam EATA in relation to expected (E).

£



Figure 3. Quadratic regression of daughter first lactation deviation on dam EATA in relation to expected (E).



Figure 4. Cubic regression of daughter first lactation deviation on dam EATA in relation to expected (E).



Figure 5. Cubic regression of daughter ME production on dam EATA in relation to expected (E).

		EATA Segments (kg)								
	≤ -350	≤-542	≤-350 -542 <	≤-158 -350<	≤34 -158<	≤ 226 34 <	≤ 418 226 <	≤ 610 418 <	>610	>418
Number of Daughters	208	20	188	1444	3659	3423	1300	257	58	315
Number of Sires	152	19	136	667	1169	1133	621	186	50	217
Dam EATA (kg)	-430	-609	-410	-232	-53	121	302	492	694	529
	(53)	(16)	(44)	(52)	(54)	(53)	(53)	(61)	(48)	(97)
Daughter First	4703	5002	4671.	4678	4872	5137	5428	5786	6000	5825
Lactation Yield (kg)	(1525)	(603)	(1330)	(1420)	(1465)	(1461)	(1569)	(1499)	(1435)	(1450)
Daughter First	6208	6567	6169	6180	6423	6754	7143	7508	7743	7551
ME Yield (kg)	(1386)	(779)	(1135)	(1361)	(1344)	(1319)	(1405)	(1400)	(798)	(1357)
Daughter First	-490	-56	-536	-336	-68	89	340	498	575	512
Dev. Yield (kg)	(1273)	(686)	(1070)	(1199)	(1206)	(1208)	(1295)	(1220)	(615)	(1150)

Table 9. Numbers of animals and yield means for EATA segments.^a

^aStandard deviations in parentheses.

Each end of the table contains an additional EATA segment which represents the two lowest segments combined or the two highest segments combined. They are not additional data but a recombination of data already in the table. They are included because of the small number of daughters in the highest and lowest EATA segments. Only four dams were in the >610 group with EATA's above 802 kg.

The distribution of daughters by size of dam EATA is depicted in Figure 6. Fewer daughters were three and four standard deviations below the EATA mean than three and four standard deviations above the mean. This was probably a consequence of culling. However, within two standard deviations of the mean there were more daughters below the mean than above the mean. In these data 3.04% of the EATA's were larger than $\mu + 2\sigma$, which is close to Syrstad's (1971) 3.05% for indexes on Norwegian Red cows. Number of daughters per sire was extremely symmetric rising from 1.2 and 1.1 for sires mated to extremely high and low EATA dams to 3.0 and 3.1 for those mated to average EATA dams.

The largest EATA's seem much too large when considering a normal distribution of EATA's. In a normally distributed population only three in one million EATA's are expected to exceed 4.5 standard deviations above the mean. In this population four in ten thousand exceeded 4.5 standard deviations. If distributed normally four in ten thousand are expected to exceed 3.3 standard deviations. Therefore, it appears $\left(\frac{3.3}{4.5}\right)^2 = .54$ of the original variation in EATA's remains in this population of EATA's. If the

largest EATA's are abnormally large because of a 46% reduction in variance attributed only to truncation selection, then 75% of the original popula-



Figure 6. Distribution of daughters by dam EATA.

tion of EATA's was saved.

Examining variances leads to a similar conclusion. Multiplying the variance of deviated daughter records by heritability of .25 gave an expected additive genetic variance of 384,400. Four times the variance of EATA's should also give the additive genetic variance in the population. However that estimate was 147,456, indicating a loss of approximately 60% of the original additive genetic variation, assuming heritability of .25. Regression, rather than selection, caused this loss of variation.

The narrow line in Figure 7 indicates the expected daughter first lactation deviated production for each EATA segment. Expected production (EP) was calculated from the equation EP = 6.8 + 1.01 (EATA-34.5). EATA was the average of dam EATA's in the group. The thick line in Figure 7 rises to actual average daughter first lactation production in each group. Production of daughters in groups with dam EATA's greater than 802 kg and less than -542 kg differed significantly (P<.05) from expected.

Within each EATA segment, correlations between dam EATA and daughter first lactation production were much lower than expected. These correlations are in Table 10. Since standard deviations of daughter production were similar in size to those for the whole population, the small correlations were due to small covariances between dam EATA and daughter yield. The two end segments were exceptions however. Daughter production of dams with EATA's over 610 kg and under -542 kg was much less variable than production in other segments. This caused high correlations between dam EATA's and daughter production in those groups. The group with EATA's less than -542 kg contained no degrees of freedom after adjusting twenty daughter records for nineteen sizes.



Figure 7. Expected and observed daughter production for EATA segments.

<u></u>					EATA	Segments	(kg)			
	≤ -350	≤-542	≤-350 -542≺	≤ -158 -350<	≤ 34 -158<	≤ 226 34<	≤ 418 226 <	≤ 610 418<	≻610	>418
Lactation Yield	07	No d.f.	11	.01	.03	.02	.01	06	.48	.04
ME Yield	10	No d.f.	11	.01	.06	•05	.00	03	.54	.12
Dev. Yield	09	No d.f.	16	02	.05	.05	.01	02	.15	.06

Table 10. Correlations between dam EATA and daughter first lactation for EATA segments.

Quadratic regressions of daughter production on dam EATA for each EATA segment are in Table 11. They vary substantially in size. The only partial regressions differing significantly (P<.05) from expected were the linear and quadratic regressions for actual 305 day daughter yield in the 35-226 kg EATA segment. Sampling errors were large.

R²'s calculated from regressing daughter deviation yields on linear and guadratic EATA as well as regressing on linear EATA alone are in Table 11. Most R^2 's were small when compared with R^2 from the unsegmented data. It is interesting daughter production is not more explainable in small pieces. This may be a consequence of adjusting for sires in each group. Each sire will remove a degree of freedom from each group he has a daughter in. Where the degrees of freedom for first lactations unsegmented was 8393, the sum of the degrees of freedom for the segments was 6360. With certain sires appearing in several groups, their worth within each group may be less accurately adjusted than in all data. This along with reduced degrees of freedom will contribute to larger error variances in each group and therefore smaller R²'s. Comparing R²'s from quadratic equations with those from linear equations showed very little additional variation in daughter deviations explained by a quadratic equation beyond a linear equation. However, in the segment of EATA less than -349 kg the quadratic equation accounted for more than twice the variation in daughter deviations accounted for by the linear equation. This equation and others from the extreme EATA segments are Figure 8. Note that the curve in the group of largest EATA's turned down as EATA increased, but there were enough daughters in the adjacent group to cause an upward turn as EATA increased when the two groups were combined.

					EATA S	egments	(kg)			
	≤-350	≤-542	≤ -350 -542 <	≤ -158 -350<	≤ 34 < -158<	≤ 226 34 <	≤ 418 226<	<u>≤</u> 610 418<	>610	→ 418
Lactation Vield				·······						
Linear	28.0	-16.0	-4.1	-6.9	.9	8.5*	10.5	1.4	463.1	-9.1
Quadratic	.033	.020	001	014	.001	.032*	016	003	342	.008
ME Yield										
Linear	32.1	-16.0	-61.9	-4,8	1.1	5.5	7.9	14.5	22.9	-1.3
Quadratic	.038	.023	069	010	003	017	012	015	010	.003
Dev. Vield										
Linear	17.8	-16.0	-15.2	-11.7	.7	3.5	6.6	-18.8	22.6	-3.4
Quadratic	.022	.023	013	023	002	010	010	.018	015	.004
R^2 (dev. vield)	.017	No d.f.	.025	.003	.002	.003	.001	.002	.024	.006
_2	000		0.01	000	000	002	000	000	000	00/
R ⁻ (lin. only)	.008		.025	.000	.002	.002	.000	.000	.023	.004

Table 11. Regressions of daughter first lactation on dam EATA within EATA segments.

*Differs significantly from expected (P<.05).



5.5

Figure 8. Quadratic regression of daughter production on dam EATA in high and low EATA segments in relation to expected (E).

Though it did not cause a cubic relationship between daughter deviation and dam EATA, daughter production in the ends of the EATA distribution deviated significantly from expected when isolated from the bulk of daughter records. There was still question as to the seriousness of these deviations, whether they were caused by a few wildly deviant daughters or a general scourge of deviancy through the tails of the distribution.

To clarify these concerns I examined information in the data on the sixty largest EATA's and sixty smallest EATA's. I ordered these by size of EATA, smallest to largest, and divided them into twenty groups with six dams per group. Groups of six were chosen to allow individual daughter fluctuations to average over a few observations but still retain enough groups in each tail of the distribution to be able to observe characteristics of the tail. Information about the low and high groups of EATA's can be compared from Table 12. The groups differed in milk and fat production, the basis of their grouping. None of the other differences were significant (P<.05), but some were interesting. Daughters of the high group milked eleven days longer than the lows. Herdmates of daughters in both groups produced more ME milk than the average population, but highs produced more than lows. Number of records for the dams and sire information on their mates were essentially the same for both groups. Dams in the high group had twice as many paternal half sisters as those in the low group. Sires with poor daughters had fewer daughters. For those interested in fat percent, please note an average of quotients is not necessarily equal to a quotient of averages.

Table 12. Information on	low sixty and high	sixty dam EATA's.
	Low 60	High 60
Dau. Days In Milk	275	286
Dau. Herdmate Milk	6708	7158
Dau. Herdmate Fat	241	257
Dau. Herdmate No.	40	40
Dau. Deviation Milk	-379	602
Dau. Deviation Fat	-9	15
Cow Recs.	5	5
Cow Milk	-1840	3016
Cow Fat	-53	90
PHS No.	584	1216
PHS Milk	-531	372
PHS Fat	-18	7
EATA Milk	-522	691
EATA Fat	-15	19
Dau. Sire Repeat. %	46	52
Dau. Sire PD Milk	24	9
Dau. Sire PD Fat	1	2

Table 13 contains dam EATA averages and daughter deviation averages for ten classes of six daughters each from the sixty lowest dam EATA's. None of the daughter deviation averages differed significantly (P < .05)from the dams' EATA averages. The high daughter averages in the lowest three EATA classes were of interest. Several possible explanations for daughter performance in these groups proved inaccurate. Since daughter deviations in Table 13 were not adjusted for sires of the daughters, they may have been large because of their sires' contributions. Of eleven daughters of dams with EATA's less than -600 kg, only five of the daughters' sires had PD's. These PD's averaged 121 kg which would account for little of the differences between daughters and dam EATA's. Another possibility was these daughters may have been first calves of their dams, and the dams were subsequently culled after one lactation. This also proved erroneous in that all dams in the lowest three groups had at least two lactations and averaged nearly five each. Two dams with EATA's less than -600 kg had two daughters each, three of which produced more milk than their herdmates.

For the lowest ten classes the regression of class daughter average, unadjusted for sires, on dam EATA average was -3.4, indicating lower EATA averages had higher daughters averages. The correlation was -.6. For a dairyman to retain a cow for five lactations with an EATA more than three standard deviations below the mean, he knows something about the cow which does not show in her EATA. Possibly she has suffered an environmental calamity which has maligned her milk production such that her EATA is no longer indicative of her genetic worth. This explanation, however, also leaves much to be desired as most relatives of these dams

Class	Mean Dam EATA	Mean Daughter Deviated Yield
1	-676	201
2	-609	-129
3	-564	5
4	-530	-741
5	-506	-734
6	-490	-386
7	-472	- 3 63
8	-465	-321
9	-459	-402
10	-454	-922
	$b_{DAU} \cdot DAM = -3.4$ $r_{DAU} \cdot DAM =6$	

Table 13. Yield means for sixty lowest EATA's, ten classes of six daughters each.

were also extremely poor milk producers.

Examining the high end of the EATA distribution I found a slightly different trend. Average production for these ten classes is in Table 14. The regression of daughter average on dam EATA average was 2.2 with a correlation of .2. This indicates daughter production was increasing more per unit of EATA increase than the average population increase. Classes twelve, fourteen, and twenty had extremely low daughter averages. It may be noteworthy that daughters of dams with EATA's averaging 843 kg produced -101 kg. Three daughters from the highest four EATA's had deviations less than -950 kg. In the highest group of six EATA's three of the mates had FD's. These three averaged 50 kg. Accounting for the value of daughter sires would lower the regression of daughter average on EATA slightly.

There seems little evidence daughters of high EATA dams produced less milk than expected. Three negative daughters from the highest four dams seems peculiar, even though they represent only .0003 of these data. One might speculate a physiological breakdown occurs when a cow reaches a high threshold of additive gene effects. This threshold response has been noted for other traits in other biological organisms. It may be informative to investigate the production of daughters of negative daughters of high EATA cows.

There is evidence daughters of the lowest EATA dams produce more milk than expected. This may be due to misevaluations of the EATA's. It seems more reasonable dams produce less milk than their genetic worth rather than daughters produce beyond their "genetic limit." However when relatives of these dams are also extremely poor producers the influence of chance on the EATA's is reduced.

Class	Mean Dam EATA	Mean Daughter Deviation Yield
11	612	622
12	621	~59
13	631	454
14	651	-619
15	660	1006
16	684	589
17	710	1208
18	736	1004
19	763	1918
20	843	-101
	^b DAU •DAM = 2.2	
	r _{DAU} · DAM = .2	

Table 14. Yield means for sixty highest EATA's, ten classes of six daughters each.

Here again an extreme accumulation of poor genes in one individual may represent an ultimate threshold where the only response is toward the mean.

We examined the accuracy of predicting daughter production from dam EATA in the extremes of the EATA distribution. Now we need to focus on the bulk of the dams where the linear regression of first lactation daughter deviation on dam EATA was larger than expected. Many researchers have reported higher heritabilities for first lactation production than later lactations. Also work has shown correlations between son's progeny test and dam's breeding value based on first lactation to be higher than those based on any other lactation. To see if a similar situation existed between dam EATA and first lactation daughter production I separated the daughter-dam pairs by number of lactations of the dam into nine groups. I then compared yield means and correlations from the various These are in Table 15. Since the EATA information included only groups. average production of the dam and number of records in the average, a dam was included in only one group. With these data it was impractical to use dam's first lactation and compare that with using an average of her first and second lactations, thereby including a dam in more than one group. That would have provided a more effective determination of the combination of dam records to be included in EATA to predict daughter production most accurately.

In groups where dams had less than three lactations in their averages their EATA's averaged less than the other group averages. With few lactations their EATA's would have been regressed more toward zero. Also they had survived fewer selection decisions than those in groups having more lactations. Average EATA in the group with more than nine records
Records	Daughters	Sires	Mean EATA	Mean Dau.ME	Mean Dau.Dev.	Correlation Dev: EATA
1	74	62	-50	6163	-28	.53
2	428	282	0	6582	-58	.08
3	1413	652	28	6653	32	.19
4	2237	833	44	6631	3	.16*
5	2145	772	42	6633	16	.18
6	1817	712	37	6616	-2	.17
7	1215	527	29	6550	-19	.21
8	615	323	41	6633	34	.20
9+	405	252	12	6688	40	.37*
A11	10349		34	6619	7	.18

*Differs significantly (P < .05) from .18 and .20.

Table 15. Yield characteristics of first lactation daughters and dams with varying numbers of dam records.

was low, but ME and deviation production of their daughters was highest of any group. Correlations between daughter first lactation deviation and dam EATA were quite variable in groups one, two, and nine where there were less than 200 degrees of freedom. Correlations were highest (.53) in the group with one dam lactation and lowest (.08) when dams had two lactations. A range that size for those two groups is unlikely considering the dam averages containing two lactations include the dams' first lactations. However, in these data the groups were independent. The only correlations significantly (P<.05) different from either .20 expected or .18 from all data differed from both. They were .16 for dams with four lactations and .37 for dams with more than eight lactations. It is interesting that groups with even numbers of dam lactations had correlations lower than eigher adjacent group. Ignoring a dam's most recent record if it is even-numbered may have merit in improving EATA accuracy.

Cubic regressions of dam EATA on within-sire daughter deviations were not different (P<.05) from linear for any group except that with only one dam record. The high correlation in that group adds to previous research the plausibility of using only first lactations to estimate breeding values of dairy cattle.

The EATA measure of a cow's genetic worth predicted daughter production without serious discrepancies. We will now turn attention to weights given various groups of relatives in EATA. Analyses in this section will be from the components equation described in the Methods section. This was a within sire multiple regression of daughter first lactation production simultaneously regressed on her dam, her dam's dam, paternal half sisters, other daughters, and maternal half sisters. Examination of these partial regression coefficients should give indication of the appropriateness of the variances and covariances supposed in calculating the EATA weights for these data.

An initial problem in this analysis was the absence of relative groups. If ignored they would count not as missing, but as relatives with average deviated milk yield of zero. If a group was absent the component for that group had to be deleted from the multiple regression equation. Therefore equations could be solved meaningfully for various combinations of available relatives as long as each independent set of equations contained the same relative groups. Numbers of daughter records with specific combinations of available relatives of the dam are in Table 16. Note the definitions of C, D, P, O, M in the table. These are relatives in the dam's EATA after one daughter was removed to be the dependent variable. The C or cow in the EATA refers to the dam of the daughter being predicted. There were no EATA's where the cow being evaluated had no records. Using first lactation daughters the most frequent combinations of EATA relatives included records of the cow, dam, and paternal half sisters. The most frequent combination (3424 of 10349) had records for all relative groups. Nearly an equal number (2941) had no daughters

					Number of L	actations	
С	omb	ina	tio	n	A11	First	
С	D	P	0	M	7516	3424	
С	D	P		м	4621	2941	
С	D	P			1636	1084	
С	D	P	0		2135	1035	
С	•	P	0		1643	614	
Ċ		P	0	м	1477	525	
C		P			573	331	
C		P		м	406	213	•
С			0		117	45	
С	D		0	M	73	34	
С					50	29	
С	D	·		M	35	26	
С	D		0		37	18	
C			0	M	31	14	
С	D		•		22	13	
с				М	5	3	
To	tal				20377	10349	

Table 16. Numbers of daughter lactations available for each combination of relatives of the dam.

^aEach of these combinations originally included an additional daughter, the daughter being predicted. C = Cow (dam of dependent daughter), D = Cow's dam, P = Cow's Paternal Half Sisters, O = Cow's daughters (excluding dependent daughter), M = Cow's Maternal Half Sisters.

.

other than the daughter being predicted which was excluded from the EATA. The most frequent groups outside those mentioned above included the cow and her paternal half sisters. Groups without the cow and her paternal half sisters numbered less than fifty first lactation daughters each. Below is a list of relatives of the dam and the percentage of daughters associated with each group.

Cow	100%
PHS	98
Dam	83
MHS	69
Daughter	55

Components equations for relative combinations which included at least the cow with an EATA and her paternal sisters would include 98% of all available first lactation daughter records in these data.

I divided the daughters constituting this 98% into eight independent groups, each group containing a specific combination of relatives in the dam's EATA. Those eight combinations are in Table 17 with numbers of sires and daughters. A dash designates a group not available. Table 17 also includes deviation milk yield averages for the five relative groups of the dam and production averages for the daughters. There are some slight changes from Table 16 in numbers of daughters in each EATA combination.

Number of daughters per sire decreased as available relative groups decreased. The more popular bulls were mated to cows with more pedigree information. Average cow deviations were large and positive while average paternal sister deviations were moderately negative. Daughter actual yield and ME yield decreased as fewer relative groups were available, but daughter yield deviated from herdmates fluctuated among groups. Therefore

	 ·		Combin	ation of	Relatives	in EATA	· · · · · · · · · · · · · · · · · · ·	······································
	CD	POM	C D	<u>P - M</u>	CD	<u>P0-</u>	<u> </u>	POM
Number of daughters	3414		2941		1031		525	
Number of sires	991		1056		501		280	
Relatives in EATA								
Cow dev. yield (kg)	286	(814)	266	(882)	276	(816)	181	(808)
Dam dev. yield	92	(914)	125	(914)	-66	(904)		
PHS dev. yield	-72	(291)	-52	(324)	-101	(327)	-81	(334)
Daus. dev. yield	69	(999)			59	(933)	-41	(830)
MHS dev. yield	-6	(828)	3	(864)			-154	(867)
Daughter actual yield	5110	(1477)	5013	(1559)	5016	(1601)	5011	(1462)
Daughter M.E. yield	6681	(1386)	6 646	(1421)	6617	(1401)	6556	(1298)
Daughter Dev. yield	2	(1234)	1	(1284)	15	(1238)	31	(1116)

Table 17. Numbers of animals and average milk yield for specific combinations of EATA relatives.^a

^aStandard deviations in parentheses.

Table 17 (continued)

Table 17 (continued)							<u> </u>	· · · · · · · · · · · · · · · · · · ·		
		Combination of Relatives in EATA								
	CD	<u>P</u>	<u> </u>	PD -	<u> </u>	<u>P - M</u>	<u> </u>	P		
Number of daughters	1084		612		213		331			
Number of sires	587		359		1.65		238			
Relatives in EATA										
Cow dev. yield (kg)	25 6	(953)	316	(870)	210	(967)	238	(978)		
Dam dev. yield	-42	(1040)								
PHS dev. yield	-66	(354)	-106	(333)	84	(395)	-64	(404)		
Daus. dev. yield			-35	(828)						
MHS dev. yield					-105	(704)				
Daughter actual yield	4967	(1497)	4958	(1448)	4902	(1342)	4860	(1550)		
Daughter M.E. yield	6589	(1409)	6495	(1315)	6450	(1496)	6366	(1538)		
Daughter dev. yield	43	(1278)	-26	(1216)	. 71	(1539)	-78	(1351)		

.

there appeared no relationship between available relative groups and herdmate production of daughters.

To evaluate the multiple regression weights obtained from the components analysis EATA weights were needed for comparison. These were computed from the EATA equations using as numbers of relatives and records averages in the data used in the multiple regression equations. It was apparent that although the average number of paternal sisters was large, most cows had only a few while some had thousands. To reduce the influence of extreme numbers on the averages a transformation was applied to the numbers (N) of animals and records. This transformation from N to N' was N' = (N+C)^P = N⁻¹ when C = 0 and p = -1. The value used for the average number of records was the reciprocal of the average of transformed variables. This was the harmonic mean. Justification for this transformation is difficult to quantitate. It seems to serve a useful purpose. Later results also substantiate its use to some degree, but one must be cautious justifying procedures with results obtained from those procedures.

Use of a harmonic mean may be defined in a generalized least squares' sense. The linear unbiased best estimator of the mean of y's where each group of y_i 's has variance V_i is the expression:

$$\frac{\Sigma \mathbf{v_i}^{-1} \mathbf{y_i}}{\Sigma \mathbf{v_i}^{-1}}$$

Where we are concerned with numbers of records, variation is a function of the numbers involved. When each N_i is weighted inversely by its contribution to the sum of the weights, an expression analogous to the estimator

above results:

$$\frac{\sum N_i^{-1}N_i}{\sum N_i^{-1}} = \frac{\sum I}{\sum N_i^{-1}} = \text{Harmonic mean of } N_i^{+s}$$

The arithmetic mean is an illogical representation of central tendency in a skewed distribution. In the EATA section variances of averages were too low when average numbers of records were used. Using harmonic means would help alleviate that problem and in this situation will give expected regression weights lower than given by arithmetic means, a correction which seems necessary. For interest I have included both arithmetic and harmonic means in Table 18.

Mean numbers of animals and records for EATA's in each combination of relative group are in Table 18. Largest discrepancies between arithmetic and harmonic means were in numbers of paternal sisters. Arithmetic means for numbers of paternal sisters were about 1000 while harmonic means were less than 28. Some sires of these EATA cows had extremely large numbers of daughters, not representative of most of these data. The only other large discrepancies between arithmetic and harmonic means were with records of the dam of the cow. Some dams had many more records than most causing differences of greater than .9 between arithmetic and harmonic means. Discussion of means from this point on will refer to harmonic means unless designated otherwise.

Cows averaged 1.7 records less when they had no daughters in their EATA's than when they had daughters. It is reasonable cows with more records have more offspring. Also if their large numbers of records

				Combinati	on of I	Relatives	in EATA	·	
Relativ	es	CDP	<u>0 M</u>	<u> </u>	- M_	<u>CDP</u>	0 -	<u>C - P O</u>	<u>M</u>
		Aª	HP	A	H	A	H	A	<u><u> </u></u>
Cow Rec	8.	5.9	5.4	4.2	3.8	5.7	5.2	6.2	5.3
Dam Rec	s.	4.7	3.5	5.0	3.8	3.0	2.1	0	
PHS No.		1130.	26.9	1102.	27.2	1015.	20.0	1039.	17.5
Rec	s.	2.7	2.4	2.3	2.1	2.6	2.3	2.6	2.4
Daus. N	0.	1.6	1.3	0		1.6	1.3	2.1	1.5
Rec	8.	2.2	1.7			2.1	1.7	2.6	2.0
MHS No.		2.1	1.6	1.9	1.5	0		1.6	1.3
Rec	8.	3.3	2.5	· 2.9	2.2			3.4	2.6
Relativ	es	CDF	·	_ C - F	<u> </u>	C - F		C - P -	
		A	H	A	H	A	Н	A	н
Cow Rec	8.	4.0	3.5	6.0	5.2	4.5	3.5	4.2	3.5
Dam Rec	8.	3.3	2.3	0		0		0	
PHS No.		895.	19.4	993.	8.1	990.	16.9	928.	9.1
Rec	S .	2.3	2.1	2.8	2.4	2.5	2.2	2.4	2.1
Daus. N	ío.	0		1.8	1.4	0		Ó	
Rec	8.			2.4	1.9				
MHS No.		0		0		1.5	1.3	0	
Rec	:8.					2.9	2.1		

Table 18. Mean numbers of animals and records in EATA's of specific relative combinations.

^aArithmetic mean.

b Harmonic mean. indicate positive value to the dairyman, then their daughters probably have been given a better than average chance to have records. Dam records averaged 1.4 records less when there were no maternal sisters than when there were maternal sisters. This is also a dam-daughter relationship.

Average numbers of paternal sisters were highest where the cow, dam and maternal sisters were also present. They were lowest where the dam and maternal sisters were not present. Less popular sires were mated to those dams. Average numbers of daughters and maternal sisters were similar from combination to combination where they were not missing.

Partial regression coefficients to predict daughter first lactation deviated milk yield from average deviations of relatives in the dam's EATA are in Table 19. They were computed from various combinations of relatives in EATA. The symbols C D P O M designate the relatives present in the EATA. A dash (-) indicates a relative not available for use in the EATA. Combinations with one or more relatives slashed (/) represent the same data included in the CDPOM combination, except slashed relatives were ignored as though they were not available for use. Therefore values corresponding to combinations with slashes were similar to those from the CDPOM combination because they included the same data. The EATA weights are values obtained when the theoretical EATA equations were solved using the average numbers of relatives and records in that combination of relatives. Standard errors of the partial regression weights are included. They were largest for paternal sister weights where the variance of paternal sister averages was low.

Regression weights from 3414 daughters with all relatives in the dam's EATA were larger than EATA weights for the cow (dam of the depen-

		Combination of relatives in	EATA		
	срром	С D P Ø M ^a	$CDP - M^{b}$		
Relative	Regr. EAT Wt. Wt	A Regr. EATA . Wt. Wt.	Regr. EATA Wt. Wt.		
Cow	.20 .16 (.032) ^c	.19 .17 (.031)	.27* .16 (.034)		
Dam	.00* .06 (.028)	•00* •06 (•028)	.02 .06 (.033)		
PHS	.21 .23 (.086)	.21 .24 (.086)	.02* .24 (.092)		
Daus.	05* .05 (.025)				
MHS	.02 .03 (.030)	.02 .03 (.034)	02 .03		
Nc. of Daughters	3414	3414	2941		

Table 19.	Partial regression coefficients to predict daughter first lactation
	deviated milk vield from relatives in her dam's EATA.

^aSlashed relative (\emptyset) indicates data were from 3414 records where all relatives were present and the slashed relative was not included in this multiple regression.

^bDash indicates that relative was not present.

^CStandard error of regression weights in parentheses.

* Differs significantly (P<.05) from EATA weight.

		Combination of rela	tives in EATA	•
Relative	<u>CDPOM^a</u>	<u>CDPO-^b</u>	<u>C Ø P O M</u>	<u>C-POM</u>
	Regr. EATA	Regr. EATA	Regr. EATA	Regr. EATA
	Wt. Wt.	Wt. Wt.	Wt. Wt.	Wt. Wt.
Cow	,20 .16	.25 .17	.20 .17	.27 .17
	(.032) ^c	(.067)).031)	(.091)
Dam	。01 .06 (.028)	.08 .05 (.059)		
Phs	,21 .22	.10 .20	.21 .22	.21 .18
	(.086)	(.163)	(.086)	(.216)
Daus.	05* .05	14* .05	05* .05	10* .07
	(.(\25)	(.058)	(.025)	(.086)
MHS			.02 .04 (.030)	07 .03 (.082)
No. of Daughters	34).4	1031	3414	525

•

٠

.

•

Table 19 (continued)

•

•

		(Combination	of rela	tives in 1	eata		
Polative	<u>CDPØM</u>		<u>CDP</u>		CPPOM		<u>C - P D -</u>	
Kelative	Regr. Wt.	EATA Wt.	Regr. Wt.	EATA Wt.	Regr. Wt.	EATA Wt.	Regr. Wt.	EATA Wt.
Cow	.19 (.031)	.17	.19 (.061)	.16	.20 (.031)	.18	.16 (.089)	.18
Dam	.01 (.028)	.06	.02 (.055)	.06				
PHS	.21 (.086)	. 24	09 (.162)	. 22	.21 (.086)	.21	.27 (.231)	.13
Daus.					05* (.025)	.05	20* (.092)	•06

.

•

78

.

.

•

· · · · · · · · · · · · · · · · · · ·		Combination of relat	ives in EATA	
N . 1. 4 for a	срром	C - P - M	Сррфи	C - P
Ketattve	Regr. EATA Wt. Wt.	Regr. EATA Wt. Wt.	Regr. EATA Wt. Wt.	Regr. EATA Wt. Wt.
Cow	.19 .18 (.031)	.13 .17 (.221)	.19 .19 (.031)	.09 .18 (.146)
Dam				
Phs	.21 .23 (.086)	.58 .20 (.540)	.21 .23 (.086)	.43 .15 (.354)
Daus.				
MHS	.02 .04 (.030)	.75* .04 (.303)		
No. of Daughters	3414	213	3414	331

Table 19 (continued)

dent daughter) and smaller for the dam and other daughters. Weights for the dam and daughters differed significantly (P<.05) from the EATA weights for the CDPOM combination. Regressions were essentially the same when daughters were deleted. As relatives other than the cow and paternal sisters were deleted by singles and pairs only the daughter regressions recained significantly low, and they were always negative. This situation will be discussed later. The cow regressions were high and the dam regressions low, but not significantly different from the EATA weights. For the 3414 daughters with CDPOM EATA's none of the regressions changed much when relatives were deleted from the analysis. The many records in the relative groups held the regression weights stable.

Regression weights from analyses where certain relatives were not available fluctuated, but standard errors were such that few regressions deviated significantly from their corresponding EATA weights. Daughters received negative regression weights significantly different from the EATA weights. Where daughters were the only missing relatives the paternal sister regression was lower and significantly different from the EATA weight. This however, was balanced by a large regression for the cow, also significantly different from the EATA weight. The only other partial regression differing from its corresponding EATA weight was a large maternal sister regression in the C-P-M analysis where there was no information on dams and daughters.

To compare meaningfully the emphasis placed on various relatives by the multiple regression equations the partial regression coefficients need be standardized for unequal variances of the independent variables.

The standardized regression coefficient β relates to the original **b** as follows.

$$\beta_{i} = b_{i} \frac{\sigma_{x_{i}}}{\sigma_{y}}$$

Table 20 includes standard partial regression coefficients for the combinations of relatives included in the multiple regression analyses. Weights with asterisks differed significantly from EATA weights before standardization.

Comparing combinations where relatives were deleted, emphasis placed on each relative essentially did not vary. The dam weight rose from .00 to .01 when any relative other than daughters was deleted. One would expect slight increase in the cow regression as maternal relatives were deleted. The weights from analyses where relatives were not available were more subject to changes since they represented independent groups of data. This makes it difficult to interpret how much of a change in a coefficient was due to compensation for a missing relative and how much was due to changes in average numbers of animals and records in each group of relatives. The cow received the most weight when only one maternal relative was missing, but she lost influence when two maternal relatives were missing. This likely was a consequence of younger cows when more than one relative was not available. Paternal sisters received 40-50% of the cow weights when dams or daughters and maternal sisters were present. However, as maternal relatives became less frequent and cows had fewer records paternal sisters received more weight, eventually more than the cow.

Combination	Cow	Dam	PHS	Daus.	MHS
CDPOM	.13	.0 0*	.05	04*	.01
СДРØМ	.13	•00*	.05		.01
CDP-M	.19*	.01	.01*		.01
СДРОИ	.13	.01	.05	04*	
CDPO-	.16	.06	.03	11*	
СрРОМ	.13		•05	04*	.01
С - РОМ	. 20		.06	07*	05
СDРØМ	.13	.01	.05		
C D P	.14	.02	02		
СрРОМ	.13		•05	04*	
C - PO -	.11		.07	14*	
Сррфи	.13		.05		.01
C - P - M	.08		.15		.34*
сррои	.13		.05		
C - P	.07		.13		

Table 20. Standard partial regression coefficients from various combinations of relatives.

*Differed significantly (P < .05) from EATA weights before standardization.

The dam received little weight except where maternal sisters alone were missing, which was a strange situation since average number of cow records was nearly unchanged while average number of dam records was 30% lower than when all relatives were present. Dams took up the slack of missing maternal sisters even though the dams had fewer records. Standard regression coefficients for daughters were large when other relatives were missing, but negative. Maternal sister weights were small except when the dam and daught .s were missing. There it received the largest weight of any relative in any combination of relatives. A small gain would be reasonable, but one this large was surprising.

It is interesting to compare standard regression weights from groups where relatives were deleted with groups where the deleted relatives were never available. There were fewer degrees of freedom and fewer relatives and records for those with missing relatives. Where there was much data and only one relative group missing the maternal relatives, especially the cow, received more weight and paternal sisters less weight than in the combinations with one relative deleted. When more than one relative group was missing and degrees of freedom reduced, the paternal half sisters and an occasional maternal relative picked up weight at the expense of the cow. This did not occur where the missing relatives had been deleted. The combinations with missing relatives were more sensitive to changes in records and relatives than were the groups where relatives were deleted. Some of the sensitivity is difficult to explain.

To determine the value of a relative in combination with various other relatives for predicting daughter milk yield the accuracies with

and without that relative must be compared. Accuracy in this case is the multiple correlation between the combination of relatives' records and independent daughters' yields. These correlations for each combination of relatives examined are in Table 21. The expected correlations were calculated as explained in the Methods section. The first two columns of correlations were based on the same 3414 daughters, deleting certain relatives actually available. Correlations in the last two columns were based on different daughters for each combination of relatives.

All correlations were lower than expected based on the 3414 daughters with all relatives available but certain of them deleted. They were about 80% of expected. A portion of this difference can be explained with a shortcoming of my multiple regression procedure. The number inserted into the regression equations for each relative group was the average deviated production of the cows in the group. It was not adjusted for numbers of animals or records in the average. The same average was used if it contained 100 animals or ten animals. Therefore the information gleaned from the components analyses corresponds to average numbers of relatives and records which were included. While this may be satisfactory for the regression coefficients it will certainly reduce the multiple correlations. More accurately described, the expected correlations are computed as though all cows in the analysis had the average numbers of relatives and records. Since the numbers of relatives and records in the production averages vary, failure to account for this will reduce the multiple correlations or, depending on your standard, increase the expected correlations. Variances and standard deviations for numbers of relatives and records where all groups were present are in Table 22.

Combination			tion ²	<u>Relativ</u> Actual	<u>e Deleted</u> Expected	<u>Relative Not</u> Actual	Available Expected	
С	D	P	0	м	.148	.192	· · ·	
С	D	P	0		.147	.192	.195	.190
С		P	0	м	.148	.188	.215	.187
С	D	P		М	.143	.188	.191	.185
С	D	P			.143	.188	.142	.180
С		P	0		.147	.185	.194	.180
C		P		м	.143	.183	.372	.176
С		₽			.143	.180	.158	.170
С					.134	.167		

Table 21. Multiple correlations between daughter first lactation deviation yield and combinations of dam relatives.

^aA space designates a relative deleted or not available.

Relative	Variance ^a	Standard Deviation
Cow Recs.	2.65	1.62
Dam Recs.	5.05	2.24
PHS No.	3623117.	1903.
Recs.	.81	.90
Daus. No.	.72	.84
Recs.	1.22	1.10
MHS No.	1.35	1.16
Recs.	2.35	1.53

 Table 22.
 Variances and standard deviations of numbers of animals

 and records in 3414 EATA's with all relative groups.

^aVariance about arithmetic mean.

These are variances about arithmetic means. Paternal sister numbers were most variable followed by numbers of dam and cow records. Not accounting for this variation will lead to multiple correlations lower than expected for average numbers of records, but I do not know how much lower.

To adjust expected correlations to account for variation of number of records in the averages the formulae would need to be adjusted to include these standard deviations. To calculate actual multiple correlations which account for varying numbers of records in the averages, the averages could be adjusted. If the weight each average would receive in its EATA is available,or if time and funds are available to calculate each of them, then one might consider multiplying each average deviation by its EATA weight and regressing daughter yield on the weighted averages. If the EATA weights are as expected, the partial regression coefficients will be one. However, there would be other things involved such as heritability and combinations of relatives and records, the consequence of which is not apparent.

Examining the multiple correlations in Table 21 where relatives were deleted indicated which relatives increased the accuracy of predicting daughter deviated production. Deleting the dam from any combination of relatives resulted in no loss of accuracy. It was expected to cause 2-4% loss in accuracy, mostly when other maternal relatives were scarce. Deleting maternal sisters in the absence of dam and daughters also did not cause the expected 2% loss in accuracy. When cow and paternal sisters were present accuracy was gained only by adding combinations of relatives which included at least daughters. Ninety-nine percent of

highest accuracy was reached with cow, paternal sisters, and daughters. This concurs with results of Deaton and McGilliard (1965). Daughters contributed accuracy even though their partial regression coefficients were negative. Deleting paternal sisters when only the cow was present caused 6% loss in accuracy, 7% expected. Accuracy fell 9%, .148 to .134 when all relatives except the cow were deleted. Thirteen percent was expected.

Multiple correlations for groups where certain relatives were not available were 3-15% larger than expected for most combinations. These are also in Table 21. There was no discernible trend to the accuracies as relative groups became less available. However, degrees of freedom were decreasing as relative groups disappeared.

From Table 21 there is a tendency to conclude most multiple correlations were lower than expected. However, all combinations where relatives were deleted represented the same degrees of freedom. Only three of the eight independent combinations yielded multiple correlations less than expected. Those three contained 50.4% of the degrees of freedom in the eight groups of data.

Table 23 contains mostly figures computed from the components analyses. They compare the partial regression coefficients and theoretical EATA weights as average index values for each combination of relatives examined. The first column contains actual mean EATA's. Columns two, three, and four contain sums of the regression weights times the average relative deviations. Sums using the partial regression coefficients were close to the actual mean EATA for the 3414 daughters with all relatives present. In groups with missing relatives the sums were variable

					Sum of	Wts. x average	e deviations	R	a TH
Com	bi	na	tio	Actual Mean EATA	Partial Regr. Wts.	EATA wts. (from har- monic means)	EATA wts. (from arith- metic means)	Partial Regr. Wts.	EATA Wts. (harmonic)
СД	P	> C		40.1	38.4	37.8	29.1	.682	,746
СD	P	, c	X		39.4	38.7	31.0	.682	.742
CD	P	, c) -	24.6	\$5.3	26.4	9.7	.700	.734
сø	F	? C	M		38.4.	35.8	27.7	.682	.726
с -	F	? C	M	5.7	46.8	40.2	30.6	.714	.722
СД	F	e ø	M		39.0	36.5	26.2	.704	.730
C D	F	? -	M	39.3	73.4	.7.8	30.8	•748	.718
СD	F	e g	i M		40.0	36.7	28.2	.704	.726
ם כ	Ē	? -	• -	28,8	53.6	23.9	13.1	.596	<u>-</u> 700
сø	Ē	? (N		38.5	39.7	31.7	.675	.718
с -	E	? () -	41.1	29.1	41.1	20.3	.505	.702
сø	Ē	? ý	6 M		39.3	34.5	25.0	.704	.708
<u>c</u> -	I	Ρ.	- м	25.1	-99.7	14.8	3.9	.962	.680
сø	İI	2	ó M		39.1	37.6	28.3	.696	.698
с -	1	Ρ.		30.0	-6.1	33.3	19.4	.628	.658

Table 23. Calculated EATA's and R 's based on mean numbers of relatives, records, and production from various combinations of relatives.

 $a_{R_{IH}} = 2(\Sigma b_{i}RHS_{i})^{\frac{1}{2}}$ where RHS_{i} is from theoretical EATA equations.

with respect to mean EATA's. This was a consequence of the sensitive weights mentioned previously. A comparison of sums using harmonic means of relative and record numbers to compute the weights with sums using arithmetic means showed the harmonic sums to be closer to the actual EATA mean in seven of eight groups. This was because average paternal sister deviations were negative and the arithmetic weights were large, reducing the size of the sums of weights times averages.

The last two columns in Table 23 compare actual R_{IH} 's from the multiple regressions with theoretical R_{IH} 's from EATA equations. Since actual genotypes of cows are not known, actual correlations between the index and daughters' genotypes cannot be calculated. I have constructed what I term actual R_{IH} 's exactly as theoretical R_{IH} 's are constructed except EATA weights were replaced with observed partial regression coefficients. This can create an unusual situation in that adding a relative group with a negative partial regression coefficient will reduce this actual R_{IH} . Possibly only absolute values of regression coefficients should be used. Another problem is that an actual R_{IH} can be larger than one, where in life that is not possible.

For groups where relatives were deleted, expected R_{IH} 's declined steadily from .746 to .698 as relatives were deleted. Actual R_{IH} 's for these groups were lower with less spread between highs and lows. Expected R_{IH} 's for groups with missing relatives were lower than those for groups with the corresponding relatives deleted. The expected R_{TH} 's

where relatives were missing also declined steadily with fewer groups present while the actual R_{IH} 's fluctuated from .505 to .962. The .962 was caused by a .75 maternal sister regression coefficient. The rank of the combinations on size of actual R_{IH} 's is nearly the same as the rank on size of actual multiple correlations.

To identify specifically causes of large and small partial regression coefficients and multiple correlations it is useful to examine standard deviations and correlations of the relatives available for EATA calculations. Table 24 contains standard deviations of average deviated production of relatives in EATA for various combinations of available relatives. Expected standard deviations were computed using repeatability = .5, standard deviation of individual deviated records = 1250 kg, and mean numbers of animals and records in each average. All standard deviations except one were smaller than expected. Most were at least 85% of expected. This would indicate 1250 kg was too large an estimate of the standard deviation. However, certain standard deviations were very close to expected. In nearly every combination of relatives the relative groups ranked the same in size of standard deviation from closest to expected to farthest from expected: daughters, maternal sisters, dam, cow, and paternal sisters. Maternal sisters in the C-P-M combination and paternal sisters in the C-PO- combination were lowest at 74% and 71% of expected standard deviations. They were about 50% as variable as expected. Standard deviations for cow averages ranged from 84% to 98% of expected, almost perfectly correlated with the average number of records in the mean average. When there were five or more records the standard

	·····		Combina	Combination of relatives in EATA							
Relative	<u>CD</u> Actual	<u>POM</u> Expected ^a	<u>CDP</u> Actual	- <u>M</u> Expected	<u>CD</u> Actual	<u>PO-</u> Expected	<u>C - P</u> Actual	0 M Expected			
Cow	814	962	882	993	816	965	808	964			
Dam	914	1002	914	99 3	904	1074					
PHS	291	368	323	369	327	386	334	394			
Daus.	999	988			933	988	830	902			
MHS	828	849	864	889			866	924			
Relative	СД	P	<u> </u>	<u>0 -</u>	<u> </u>	<u>P – M</u>	<u> </u>	<u>P </u>			
<u></u>	Actual	Expected	Actual	Expected	Actual	Expected	Actual	Expected			
Cow	953	1002	870	965	967	1002	978	1002			
Dam	1040	1059									
PHS	3.54	390	333	471	395	399 .	404	462			
Daus.			827	938							
MHS					704	954					

Table 24. Standard deviations of average deviated production of relatives in EATA for combinations of relatives in EATA.

^aRepeatability = .5, Standard deviation of individual records = 1250 kg.

deviations were 85% of expected, while with 3.5 records they were 97% of expected. This indicates repeatability probably declines with additional records. To equalize the expected and actual standard deviations for 5.4 records, repeatability would be .30, and for 3.5 records..44. Correlations between adjacent records on cows may be higher than between nonadjacent records as reported by Butcher and Freeman (1968). Selection may have reduced the variation in averages of older cows. Daughters, maternal sisters, and dams were closest to expected standard deviations. Paternal sisters were least variable as expected because of the large numbers of records in their averages. They were also farthest from expected. Across relative combinations, standard deviations of a specific relative ranked by size about the same as the corresponding expected standard deviations.

Correlations between daughter first lactation deviated milk yield and average deviations of relatives in her dam's EATA are in Table 25. Variation in daughter records contributed by the daughters' sires has been removed. Therefore an only daughter of a sire contributed no variation. Most correlations were lower than expected, indicating the covariances were low since the variation in averages of relatives was also lower than expected. The daughters were always negatively correlated with the other daughters, their maternal half sisters. This was a consequence of regressing each daughter on all remaining daughters. It can be explained effectively by an example. Consider the daughter average when daughters are withdrawn as dependent variables one at a time with replacement. When the dependent daughter is a high producer the daughter

		Ço	mbinat:[0]	n of rela	tives in	EATA		
Relative	<u>CDP</u> Actual	<u>O M</u> Expected ^a	<u>CDP</u> Actual 1	- <u>M</u> Expected	<u>CDP</u> Actual	0 - Expected	<u>C - P</u> Actual 1	<u>0 M</u> Expected
Cow ^b	.134	.167	.190	.162	.156	.167	.186	.167
Dam	•022 [.] *	.081	.046	.082	.067	.075		
PHS	•067*	.109	.044*	.109	.043	.104	.098	.102
Daus.	015*	.082			070	.082	045*	.090
MHS	.021	.048	.006	.045			020	.043
S.E. ^C	.020		.023		•044		.064	

Table 25. Correlations between daughter first lactation deviated yield and average deviations of relatives in her dam's EATA.

^aRepeatability = .5, Heritability = .25.

^bCow in the EATA is the dam of the dependent daughter.

^CStandard error = $1/(N-3)^{\frac{1}{2}}$ where N = (no. daughters-no. sires absorbed). *Differs significantly (P<.05) from expected.

Tai	h1,	ρĵ	25.	- (' con	ti	nu	eď)
	U	- 4	~ ~ •	· · ·		~+			,

•

•

Relativ	ve <u>CDP</u> Actual E	cpected	<u> </u>	<u>0 -</u> Expected	<u>C - P</u> Actual 1	- <u>M</u> Expected	<u> </u>	Expected
Cow	.139	.161	.118	.167	.046	.161		
Dam	.034	.076						
PHS	007*	.103	.098	.086	.138	.101	.144	.087
Daus.			129*	.086				
MHS					.334*	.042		
S.E.	.045		.063		.149		.105	

average falls without her. When she is a poor producer it rises. This creates a negative correlation between the dependent daughters and the remaining independent daughter averages. This correlation cannot be less than $\frac{-1}{m}$ where m is the number of daughters remaining in the average.

For 1.7 daughters $\frac{-1}{m} = -.6$. This would occur if there were no variation in averages of unrelated groups of daughters, who are maternal half sis-

ters within each group. The variance of group averages would be,

$$\left(\frac{1+(n-1)\rho}{n}\right)\sigma^2 = 0 \quad \text{minimum}$$
$$\rho = \frac{-1}{n-1}$$

where n = m+1, σ^2 is the variance of individual records, and ρ is the correlation between daughters within each group. The situation with these data is not that serious, but it appears differences within groups of related daughters are larger than between unrelated groups.

These negative covariances between related daughters caused negative regression coefficients for daughters to predict another daughter. However, negative regressions contribute positively to \mathbb{R}^2 . It is questionable, though, whether that contribution to \mathbb{R}^2 was the same as it would have been had the negative correlation not been built into the analysis. To avoid the problem only one daughter per dam should be used as a dependent variable.

Correlations between the daughter and the cow (daughter's dam) were generally lower than expected though not significantly (P<.05) different from expected. The .134 correlation where all relatives were

present compares favorably with the .14 reported by Deaton and McGilliard (1965). When the cows had only one daughter each, the dependent one, the correlation with those daughters was 17% higher than expected (.190, .162). There appears to be a positive correlation between permanent environmental effects of these cows and their daughters' first lactation deviated yield.

Dams' (grandams of the daughters) correlations with daughters' yield were relatively lower than those of the cow which may indicate less permanent environmental correlation between dam and daughter than cow and daughter. Dams were older than cows and more removed from daughters in the years they were milking. Only the dam correlation where all relatives were present was lower and significantly (P<.05) different from expected.

Paternal sister correlations with daughter production were low except where the average number of paternal sisters was small. When there are few paternal sisters a larger percentage of them is likely to be in the same herd than when there are many. This would increase environmental correlations between daughters and paternal sisters.

Correlations between average deviated production of various relatives in EATA are in Table 26. Nearly all correlations were lower than expected, though not all significantly different from expected. Flock (1964) reported predicted correlations about 10% higher than actual correlations. There are a number of possible explanations for the expected correlations being too large. Heritability of .25 may be too high or less probably repeatability of .5 may be too low. A reduction in heritability will reduce the expected correlations proportionately. A

			Combina	tion of	relativ	<u>es in E</u>	ATA	
Rolatives	<u>C D P</u>	<u>OM</u>	<u>CD P</u>	<u>– M</u>	<u>C D P</u>	0 -	<u>C - E</u>	<u>N O M</u>
	Act.	Exp. a	Act.	Exp.	Act.	Exp.	Act.	Exp.
Cow, Dam	. 15 7 *	.202	.176	.198	.071*	.188		
Cow, PHS	.142*	.236	.216*	.266	.130*	.262	.224	.257
Cow,Daus.	.149*	. 205			.168	. 205	.145	. 225
Cow, MHS	.051*	.120	.105	.111			.168	.110
Dam, PHS	045*	0	004	0	008	0		
Dam, Daus.	.042*	.099	·		.051	.092		
Dam, MHS	.172*	. 230	.168*	.221				
D.112 D	05.5.4	10/			000+	1.00	050	107
PHS, Daus.	•055*	.134			•039*	.128	.050	13/
PHS, MHS	002	0	.043	0	•		.057	• 0
Daus,MHS	.026	•058					.012	.058
S.E. ^b	.020		.023		.044		.064	

Table 26. Correlations between average deviated yield of relatives in EATA.

^aRepeatability = .5, Heritability = .25

^bStandard error = $1/(N-3)^{\frac{1}{2}}$ where N = (no. daughters - no. sires absorbed).

*Differs significantly (P<.05) from expected.

			Combinat	ion of	relative	s in EA	TA	· · · · · · · · · · · · · · · · · · ·
	CDP	• - _	<u>C - P</u>	0 -	<u>C - P</u>	<u>- M</u>	<u>C - P</u>	
	Act. 1	Exp.	Act.	Exp.	Act.	Exp.	Act.	Exp.
C ow, Dam	.132	.184						
Cow, PHS	.124*	• 250 ·	. 203	. 215	059*	• 244	.218	.211
Cow, Daus.			.075*	.216				
Cow, MHS					073	.102		
Dam, PHS	.015	0						
Dam, Daus.								
Dam, MHS								
PHS, Daus.			034*	.110				
PHS, MHS					020	0		
Daus.,MHS								
S.E.	.045		.063		.149		.105	

Table 26. (continued)

situation not accounted for in the expected correlations is a negative correlation between records initiated in the same herd, year, and season. When a record is large other records tend to be small because they are deviated from a herdmate average which includes the large record. The deviations will be negatively correlated, but only slightly because herd-year-seasons for EATA calculations are required to have at least sixteen records or cows from an additional herd-year-season are included in the herdmate average. Selection of cows which produce more milk than average will reduce genetic variation and expected covariances between relatives. Expected correlations in Table 26 have not been adjusted for this.

Within the combination where all relatives were present nearly all correlations between averages of relative groups were smaller than expected and differed significantly (P<.05) from expected. Parent-offspring correlations (cow-dam, cow-daus, dam-MHS) were closest to expected within most combinations of available relatives. They were likely to have slightly correlated permanent environmental effects. If genetic worth of dams was correlated with genetic worth of their mates for milk yield then the covariance between cows and their dams would increase. In these data, however, the correlation between dam EATA and mate PD was .02, so this should have little effect. When there were no maternal sisters and the cow had more than one daughter the cow-dam correlation was only 38% of expected. The dams in this group were poor producers (-66) and the cows high producers (276). The fact that the cows remained after all maternal sisters departed or never arrived may explain the low correlation between the cows and their dams in terms of milk yield.
Cow-paternal sister correlations were lowest when there were many paternal sisters. With small variances, the reverse of this would seem probable. However, it appears when there are fewer paternal sisters they are probably concentrated more per herd in fewer herds. There would probably be additional covariation between their permanent environments. Some may also be full sisters which would contribute to the expected covariance an additional 25% additive genetic and dominance variation as well as epistatic variation.

Daughters had low correlations with dams, maternal sisters, and paternal sisters. Some of this would have been caused by removing from and replacing daughters in the daughter average. If a cow had more than one daughter, the daughter average would be different each time she appeared in these data. Averages for other relatives remained constant so the covariances between the fluctuating daughter averages and other relatives would be reduced.

The discussion of standard deviations and correlations was tedious. A brief summary may be worthwhile. Standard deviations were lower than expected but most at least 85% of expected. Causes may be too large an estimate (1250 kg) of standard deviation for individual records, not accounting for selection, variable repeatability, and varying numbers of records per animal when calculating variances of averages. Correlations and covariances were low, possibly due to not accounting for selection, heritability lower than that used, negative correlations between related daughters, and negative correlations between records from the same herdyear-season. Correlations nearest expected were between relatives likely to have positively correlated permanent environments, and half sisters

who actually may have been full sisters.

This project has discussed many aspects of using EATA to predict daughter first lactation deviated milk yield. Since most dams and daughters produce in the same herd, these results are applicable basically to within herd selection, which the dairyman practices. Another aspect of the problem is the ability to identify cows with highest genotypes for milk yield from among cows from different herds to produce sires for artificial insemination use. For EATA to fulfill this purpose most effectively, each cow must be evaluated with respect to the same genetic base or the same environmental base. Deviating records standardized for days in milk, frequency of milking, age, and season of calving from their regressed adjusted herdmate averages was intended to remove environmental variation from differences between records. It deviates the record from the herdmate average and credits that deviation with an additional $\left(1-\frac{N-1}{N+1}\right)X$ 100% of its difference from breed average as an ad-

justment for numbers of records in the herd average. This method does not account for genetic differences between herds. It removes from the record what is common to cows in that herd-year-season. Therefore genetic differences between herd-year-seasons are also removed. However, this is part of what we wish to measure with the cow index. The cow should get credit for competing against cows of higher or lower than average genetic worth.

To evaluate the genetic worth of specific herds, a genetic comparison is needed. Paternal half sisters milking in different herds provide a practical base of comparison of one herd to another. Using at least

these sisters to compute an average breeding value for the cows in a herd should give an estimate of the genetic worth of that herd compared with other herds. This can be used to index cows such that cows from different herds may be compared on the same genetic basis.

The problem of meaningfully comparing cows from different herds has created much interest with respect to choosing dams of future sires and buying cows. There needs to be research into methods of accomplishing this as well as research into whether changes would be worthwhile. The problem of comparing cows to the same genetic base is mostly a problem of measuring what we intend to measure with standardized deviated milk records, measure production from a standard genetic base. One could compensate with the index or refine the measure.

Table 27 summarizes the fractions of daughters with first lactation ME milk production greater than herdmate production for various groups of dam EATA and sire FD. The correlation between sire FD and dam EATA for these data was .02. Standard deviation for EATA was 192 kg and for FD was 264 kg. The marginal fractions of daughter yield greater than herdmates for sires increased almost linearly from .26 for sires with FD's less than μ -2 σ to .76 for sires with FD's greater than μ + 2 σ . However, these records were included in the FD's. Marginal fractions for daughters grouped by dam EATA were also quite linear except for small numbers in groups less than μ -3 σ and greater than μ +4 σ . To compare with sires, the fraction for EATA's less than μ -2 σ was .38 and more than μ + 2 σ .68. A second margin for dam EATA's includes all first lactation daughters in these data, not restricted to daughters of sires with FD's.

	Sire PD							A11
Dam EATA	<u>≤</u> -2σ	-2σ, -σ	-σ, μ	μ, σ	σ, 2σ	>2ơ		Sires
>4o	0		0	0	0	0	0	.25
3σ, 4σ	ο		+	+	+	ο	.81	.76
2σ , 3 σ	ο	+	+	+	+	+	.65	.66
σ, 2σ	+	+	+	+	+	+	.66	.63
μ, σ				+	+	+	.56	.54
-σ, μ				+	+	+	.50	.48
-2ơ, -ơ					+	+	.40	.40
-3ơ, -2ơ					+	+	.36	.34
<u>≤</u> -3σ	ο	+	+	+		ο	.64	.50
	.26	.43	.48	.57	.66	.76	.53	.51

Table 27.	Fractions of daughters with first lactation ME yield	greater						
	than herdmates. ^a							

^a μ , σ designates μ <EATA $\leq \mu$ + σ . e.g. (μ , μ + σ].

$$\sigma_{\rm EATA} = 192 \, \rm kg, \ \mu_{\rm EATA} = 34 \, \rm kg.$$

$$\sigma_{\rm PD} = 265 \ \rm kg, \ \mu_{\rm PD} = 64 \ \rm kg.$$

Plus indicates more than 50% of the daughters in that cell produced more milk than their herdmates. Blank indicates 50% or less produced more milk than herdmates. Zero indicates no daughters in that cell.

In the body of Table 27 plusses indicate more than 50% of the daughters in that cell produced more milk than herdmates. Blanks indicate 50% or less daughters produced more than herdmates, and zeros designate cells which contained no daughters. Plusses were concentrated in the upper triangular portion of the grid where dam EATA's or sire PD's were above average. Plusses in the lower right and blanks in the upper left seem to indicate higher probability of an above average daughter from a mating of high sire, low dam than low sire, high dam. However, there were few daughters on the bottom line and in the upper left of the table.

SUMMARY

Correlation of dam EATA with daughter first lactation deviated milk yield was .18 where .20 was expected. Dam EATA was also positively correlated with ME production of daughter herdmates. This was caused by some herds culling cows more intensely than other herds or using a higher proportion of genetically superior sires than other herds.

Linear regression of dam IATA on daughter first lactation deviated yield was 1.15 with a standard deviation of .07. The regression was significantly larger than the expected regression of 1.0. Any automaticity in the analysis would have inflated the observed regression. Quadratic and cubic regressions were not significant, but closer examination revealed daughters of the lowest EATA dams produced more milk than expected.

The distribution of EATA's was skewed right with a mean of 34 kg and standard deviation of 192 kg. The variance of EATA's was less than expected.

Daughter first lactation deviated yield was regressed simultaneously on groups of relatives included in the dam EATA's. Combinations of relatives in the dam EATA were required to contain at least the cow (dam of the daughter) and her paternal sisters. This included 95% of the daughters. One third of the daughters had all five relative groups in their dams' EATA's. Partial regression coefficients for the daughters of the cow were negative due to removing daughters from the daughter average one at a time as dependent variables and replacing them. Cow weights were larger than expected and dam weights were smaller than expected.

Using the daughters whose dams' EATA's contained information on all relative groups, relatives were deleted one and two at a time. Emphasis placed on each relative group essentially did not vary as groups were deleted.

In combinations where relatives were missing, standard regression weights were more sensitive to changes in available relatives. Certain relatives took up the slack of missing relatives, though not always reasonably. Paternal sisters commanded more weight than the cows when maternal relatives were missing. Cows received the most weight otherwise.

Multiple correlations of daughter first lactation deviations with relatives in their dams' EATA's were probably reduced somewhat by not weighting averages of relatives by the numbers of records in the averages. Multiple correlations for three of the eight combinations of relatives representing independent data were smaller than expected. These accounted for 50.4% of the degrees of freedom in the eight groups. The correlation for the combination with all relatives present was lower than expected due to a low correlation between daughter yield and cow average in the EATA. This low correlation persisted for all combinations where relatives were deleted because they represented the same data. Dams and maternal sisters added little accuracy to predicting daughter yield. Most correlations for combinations with missing relatives were larger than expected but variable.

Standard deviations of relative averages in the EATA's were smaller than expected as the variance used to calculate the expected standard deviations was probably low. Correlations between relatives were also

smaller than expected. Factors possibly contributing to misevaluating expected variances and correlations were selection, environmental correlations, heritability estimate too large, repeatability not accounting for numbers of records, negative correlations between dependent daughter and average of remaining daughters, negative correlation between records made in the same herd-year-season, and calculating variances of means using average records per animal rather than individual records per animal.

EATA seems to be fulfilling most of its potential for predicting daughter milk production where dams and daughters produce in the same herd. There is wide interest in comparing EATA's of cows in different herds to choose dams to save bulls from. The accuracy of these comparisons of dams from different herds merits some research to determine the extent of the problem and efficient solutions if necessary.

Further research might be useful to determine information which ought to be included in a selection index for milk production. This research should investigate not only relatives, such as the maternal grandsire, but also indicator traits such as percent fat or physiological measures. One must be cautious about excluding information which does not add to the accuracy of the index. It may contribute to the popularity of the index for the dairyman, leading to genetic progress.

REFERENCES

- Bereskin, B. 1963. Effects of genetic and environmental variance on dairy sire evaluation. Ph.D. Thesis. Iowa State University, Ames, Iowa.
- Bereskin, B., and A. E. Freeman. 1965. Genetic and environmental factors in dairy sire evaluation. II. Uses and limitations of deviation records and the role of dams. J. Dairy Sci. 48:352.
- Butcher, D. F. and A. E. Freeman. 1968. Heritabilities and repeatabilities of milk and milk fat production by lactations. J. Dairy Sci. 51:1387.
- Butcher, K. R. 1973. Estimating son's progeny test for milk yield from information on his sire, dam, and maternal grandsire. Ph.D. Thesis. North Carolina State University, Raleigh, North Carolina.
- Deaton, O. W., and L. D. McGilliard. 1964. First, second, and third records of a cow to estimate superiority of her daughters. J. Dairy Sci. 47:1004.
- Deaton, O. W., and L. D. McGilliard. 1965. Weighting information from relatives to select for milk in Holsteins. J. Dairy Sci. 48:365.
- Eastwood, B. R. 1968. Lactation summary, herd ranking, and preliminary sire summary procedures. Unpublished mimeographed paper. Ames, Iowa, Iowa State University, Dairy Extension, Bulletin DyS-961.
- Flock, D. K. 1964. Selection among Holstein-Friesian cows. Ph.D. Thesis. Iowa State University, Ames, Iowa.
- Freeman, M. G. 1970. What has been realized from pedigree selection of dairy bulls. Unpublished mimeographed paper presented at American Dairy Science Association meeting, Gainesville, Florida. Guelph, Ontario, University of Guelph, Department of Animal Science.
- Hazel, L. N. 1943. The genetic basis for constructing selection indexes. Genetics 28:476.
- Henderson, C. R. 1963. Selection index and expected genetic advance. Conference on statistical genetics and plant breeding. National Academy of Sciences-National Research Council 982:141.
- Lush, J. L. 1947. Family merit and individual merit as bases for selection. The American Naturalist 81:241, 362.
- Lush, J. L. 1960. Improving dairy cattle by breeding. I. Current status and outlook. J. Dairy Sci. 63:702.

- Miller, R. H. 1968. Estimating breeding values of cows. Proceedings of the National Extension Seminar on the Effective Use of Breeding Values of Dairy Cows and Sires for Production Traits, Madison, Wisconsin. Madison, Wisconsin, University of Wisconsin.
- Molinuevo, H. A., and J. L. Lush. 1964. Reliability of first, second, and third records for estimating the breeding value of dairy cows. J. Dairy Sci. 47:890.
- Morillo, F. J., and J. E. Legates. 1970. Evaluation of progeny tests of dairy sires in single herds. J. Dairy Sco. 53:1069.
- Plowman, R. D. and B. T. McDaniel. 1968. Changes in USDA sire summary procedures. J. Dairy Sci. 51:306.
- Robertson, A., and J. M. Rendel. 1950. The use of progeny testing with artificial insemination in dairy cattle. J. Genetics 50:21.
- Skjervold, H. 1963. The optimum size of progeny groups and optimum use of young bulls in A.I. breeding. Acta Agriculturae Scandiπavica 13:131.
- Skjervold, H., and A. K. Odegard. 1959. Estimation of breeding value on the basis of the individual's own phenotype and ancestors' merits. Acta Agriculturae Scandinavica 9:341.
- Smith, H. F. 1936. A discriminant function for plant selection. Ann. Eugenics 7:240.
- Syrstad, O. 1971. Selection of cows for planned matings. Acta Agriculturae Scandinavica 21:50.
- Thomson, G. M., and A. E. Freeman. 1970. Environmental correlations in pedigree estimates of breeding value. J. Dairy Sci. 53:1259.
- VanVleck, L. D. 1969. Relative selection efficiency in retrospect of selected young sires. J. Dairy Sci. 52:768.
- VanVleck, L. D., and G. E. Bradford. 1966. Genetic and maternal influence on the first three lactations of Holstein cows. J. Dairy Sci. 49:45.
- VanVleck, L. D., and H. W. Carter. 1972. Comparison of estimated daughter superiority from pedigree records with daughter evaluation. J. Dairy Sci. 55:214.
- Vinson, W. E. 1971. The use of independent culling levels and selection index procedures in selecting future sires for artificial insemination. Ph.D. Thesis. Iowa State University, Ames, Iowa.

- Walton, R. E. 1970. Designing A.I. programs for optimum genetic progress. Unpublished mimeographed paper presented at American Dairy Science Association meeting, Gainesville, Florida. Deforest, Wisconsin, American Breeders Service.
- Young, S. Y. 1961. The use of sire's and dam's records in animal selection. Heredity 16:91.

ACKNOWLEDGEMENTS

I sincerely appreciate the direction and help offered by Dr. A. E. Freeman and the analytical questioning from Dr. J. L. Lush.

Mostly I appreciate Jan.