

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

---

Faculty Papers and Publications in Animal  
Science

Animal Science Department

---

June 1970

## Misidentification and Sire Evaluation

L. Dale Van Vleck

University of Nebraska-Lincoln, [dvan-vleck1@unl.edu](mailto:dvan-vleck1@unl.edu)

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



Part of the [Animal Sciences Commons](#)

---

Van Vleck, L. Dale, "Misidentification and Sire Evaluation" (1970). *Faculty Papers and Publications in Animal Science*. 427.

<https://digitalcommons.unl.edu/animalscifacpub/427>

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

# Misidentification and Sire Evaluation

L. D. VAN VLECK

Department of Animal Science, Cornell University  
Ithaca, New York 14850

## Abstract

Biases in evaluation of sires result from misidentified records in the sire group averages used for sire evaluation. The bias increases as the fraction of misidentified cows increases. Use of variance components estimated from misidentified records also contributes to the bias. Estimates of the correlation between estimated and true genetic value, as well as estimates of genetic progress due to selection, are biased when misidentified records are included in sire group averages.

## Introduction

An earlier paper (1) reported the effect of misidentification on estimation of components of variance for sire and within sire. The bias in estimation of heritability from 4 times the intrasire correlation was approximately proportional to  $1-p$  where  $p$  is the fraction of cows correctly identified in each sire group. The results were nearly the same for 4 different patterns of misidentification.

A related problem with misidentified records is in evaluation of sires from averages of cows which are identified as being by a particular sire when some of the cows in the group are by other sires. The object of this paper is to demonstrate the biases in sire evaluation and in prediction of genetic progress when misidentified records are used in sire evaluation and in estimation of variance components.

## Statistical Considerations and Results

The models considering the effect of misidentified records on sire evaluation are 2 of those considered earlier in describing biases in estimation of components of variance (1) because the results for those 4 models were similar for estimating sire and within sire components of variance.

*Model 1.* All misidentified cows in a sire group have the same sire:

$$\bar{y}_i = \mu + n_{ii}s_i + (n_i - n_{ii})s_{ik} + \bar{e}_i,$$

where  $\bar{y}_i$  is the average of "daughters" of the  $i^{th}$  sire,

$\mu$  is a constant common to all records,

$s_i$  is a random effect common to all daughters of the  $i^{th}$  sire,

$s_{ik}$  is a random effect common to all cows incorrectly identified as having sire  $i$  when the real sire is sire  $ik$ ,

$\bar{e}_i$  is the average of uncorrelated random effects common to records of the  $n_i$  cows in the average,

$n_{ii}$  is the number of correctly identified daughters, and

$n_i - n_{ii}$  is the number of incorrectly identified cows in the average.

The  $s_i$  and  $s_{ik}$  are uncorrelated with mean, zero, and variance,  $\sigma_s^2$ . The  $e_{ij}$  are uncorrelated with mean, zero, and variance,  $\sigma_e^2$ .

*Model 2.* Each of the misidentified cows in a sire group has a different sire:

$$\bar{y}_i = \mu + n_{ii}s_i + \sum_{k=1}^{n_i - n_{ii}} s_{ik} + \bar{e}_i.$$

In these models the records have been adjusted for fixed factors and expressed as deviations from their appropriate population means.

If each cow has one record, the regression factor for predicting the daughter superiority or one-half the sire's genetic value ( $s_i = g_i/2$ ) is for Model 1:

$$b = \frac{\text{Cov}(\bar{y}_i, s_i)}{\text{Var}(\bar{y}_i)} = \frac{(n_{ii}/n_i)\sigma_s^2}{[2n_{ii}(n_{ii} - n_i)/n_i^2 + 1]\sigma_s^2 + \sigma_e^2/n_i}.$$

For Model 2, the regression coefficient is:

$$b = \frac{(n_{ii}/n_i)\sigma_s^2}{[(n_{ii}^2 + n_i - n_{ii})/n_i^2]\sigma_s^2 + \sigma_e^2/n_i}.$$

These differ from the usual regression coefficient when all cows are correctly identified:

$$b = \sigma_s^2/(\sigma_s^2 + \sigma_e^2/n_i).$$

Four possibilities, other than models, will be considered in sire evaluation with misidentified records. The data to estimate  $\sigma_s^2$  and  $\sigma_e^2$  can be correctly identified or partly misidentified. The records in the group averages for evaluation can be correctly identified or partly misidentified. Situation 1 is correct identification in estimation of variance components and of cows for sire evaluation. Situation 2 is the use of unbiased estimates of variance compo-

nents in sire evaluation when a fraction,  $1-p$ , of the cows are known to be misidentified. This is the appropriate procedure for evaluation of sires with some misidentified cows in the sire group averages. Situation 3 is the use of variance components estimated from misidentified data in sire evaluation when a fraction,  $1-p$ , of the cows are misidentified. Correct identification is, however, incorrectly assumed both in estimation of variance components and in sire evaluation. This situation is the one that ordinarily occurs. Situation 4 is equivalent to Situation 2 except estimates of variance components are biased because of misidentified records which are assumed incorrectly to be correctly identified. These biased estimates of variance components are then used in evaluation of sires which are known to have a fraction,  $1-p$ , of incorrectly identified cows with records in the sire group averages.

Since there is a large number of combinations of proportions misidentified in estimation and in evaluation, the proportion misidentified will be assumed to be the same for records in estimating variance components as in evaluation for Situation 4. The number of cows in each sire group is also variable in the estimation of the variance components. The previous report (1) showed this number had little effect on the bias in estimation. Calculations for group size of 20 in estimation of variance components are reported since the results are similar for all sizes.

Formulas for expected values of the regression coefficients are in Table 1 for the 4 situations of correct and incorrect identification if

$p$  is the proportion of cows correctly identified in a sire group, i.e.,  $p = n_{11}/n_1$ . A further assumption is that  $p$  is the same for each sire group in the records used to estimate variance components.

Comparison of these regression coefficients for specific numbers of animals with records in a group will show what proportional bias in estimation of genetic value is introduced because of misidentification in the group of records for sire evaluation. The regression coefficients are in Tables 2, 3, and 4 for the intrasire correlation,  $t = \sigma_s^2/(\sigma_s^2 + \sigma_e^2) = .04, .06$ , and  $.08$  which correspond to heritabilities of  $.16, .24$ , and  $.32$ . Calculations were done for other combinations of numbers per group, fraction correctly identified, and heritability, but those reported are of greatest interest. Only the regression values for Model 1 of misidentification are reported since they agree closely with the values for Model 2. The regression coefficients in the tables show the bias in evaluation since the correct regression coefficients are the Situation 2 values. For example, the procedures using biased estimates of variances do not regress the daughter averages enough as can be seen from comparing the lines for Situations 3 and 4 with the lines for Situation 2.

The correlations between the true daughter superiority and the estimates derived from the 4 regression coefficients appear different since the regression coefficients are different, but the correlations are the same since the ranking of all bulls with the same number per group is the same no matter which regression

TABLE 1. Expected values of apparent weighting factors for group averages in sire evaluation for intrasire correlation,  $t$ .

Situation	Bias in variance estimates	Fraction misidentified in evaluation	Model 1	Model 2
1	No	None	$t/[t + (1-t)/n_s]$	$t/[t + (1-t)/n_s]$
2	No	Known fraction	$pt/[xt + (1-t)/n_s]$	$pt/[(z_i t + 1-t)/n_s]$
3	Yes	Assume none incorrectly	$wt/[wt + \{[n_v(1-x)/(n_v-1)]t + 1-t\}/n_s]$ $ut/[ut + \{[(n_v-z_v)/(n_v-1)]t + 1-t\}/n_s]$	
4	Yes	Known fraction	$pwt/[xwt + \{[n_v(1-x)/(n_v-1)]t + 1-t\}/n_s]$ $put/[ \{z_i ut + [(n_v-z_v)/(n_v-1)]t + 1-t\}/n_s]$	

$$t = (\sigma_s^2)/(\sigma_s^2 + \sigma_e^2).$$

$p$  = Proportion of correctly identified cows in sire group.

$$x = 2p^2 - 2p + 1.$$

$$z_i = p^2 n_i + p - 1 \quad (i = v \text{ or } s).$$

$$w = x - (1-x)/(n_v-1).$$

$$u = (z_v - 1)/(n_v - 1).$$

$n_v$  = number of cows in each group in estimation of variance components.

$n_s$  = number of cows in each group in sire evaluation.

$$\text{coefficients are used, i.e., } r_{GG} = \frac{\text{Cov}(b \bar{y}_i, s_i) / \sqrt{\text{Var}(b \bar{y}_i) \text{Var}(s_i)}}{\text{Cov}(\bar{y}_i, s_i) / \sqrt{\text{Var}(\bar{y}_i) \text{Var}(s_i)}}$$

for all  $b$ . Therefore, the correlations between true and estimated value are the same for all evaluation procedures using the same records. Expected genetic progress is also the same for

any of these regression equations with equal numbers in each sire group.

The usual way to estimate genetic progress per generation from selection can be written:  $\Delta G = b \sigma_{\bar{y}} D$  where  $b$  is the regression coefficient in sire evaluation,  $\sigma_{\bar{y}}^2$  is the variance of the sire group average, and  $D$  is the standard selection intensity factor based on either the

TABLE 2. Apparent weighting factors for group averages in sire evaluation with various combinations of misidentification for intrasire correlation = .04.

Situation	Bias <sup>a</sup>	Fraction correctly identified	Number of cows per group in evaluation						
			10	20	30	50	70	100	1,000
1	No	1.00	.294	.455	.556	.676	.745	.806	.977
2	No	.80	.260	.426	.541	.690	.782	.870	1.204
3	Yes	.80 <sup>b</sup>	.214	.353	.450	.577	.656	.732	.963
4	Yes	.80	.184	.318	.420	.566	.664	.764	1.179
2	No	.85	.270	.437	.550	.694	.781	.863	1.138
3	Yes	.85 <sup>b</sup>	.232	.376	.475	.601	.678	.751	.967
4	Yes	.85	.209	.354	.459	.603	.697	.789	1.123
2	No	.90	.280	.446	.556	.692	.774	.849	1.079
3	Yes	.90 <sup>b</sup>	.251	.402	.502	.627	.702	.771	.971
4	Yes	.90	.237	.390	.497	.636	.723	.805	1.071
2	No	.95	.287	.451	.557	.686	.761	.830	1.025
3	Yes	.95 <sup>b</sup>	.272	.428	.528	.651	.723	.789	.974
4	Yes	.95	.265	.423	.528	.659	.738	.810	1.022

<sup>a</sup> Bias in variance estimates.

<sup>b</sup> Incorrectly assume all correctly identified.

TABLE 3. Apparent weighting factors for group averages in sire evaluation with various combinations of misidentification for intrasire correlation = .06.

Situation	Bias <sup>a</sup>	Fraction correctly identified	Number of cows per group in evaluation						
			10	20	30	50	70	100	1,000
1	No	1.00	.390	.561	.657	.761	.817	.865	.985
2	No	.80	.356	.547	.665	.805	.885	.956	1.220
3	Yes	.80 <sup>b</sup>	.293	.453	.554	.674	.744	.806	.975
4	Yes	.80	.259	.424	.539	.688	.781	.868	1.302
2	No	.85	.368	.556	.671	.803	.877	.943	1.151
3	Yes	.85 <sup>b</sup>	.315	.479	.579	.679	.763	.821	.978
4	Yes	.85	.291	.463	.578	.720	.805	.883	1.141
2	No	.90	.377	.561	.671	.794	.862	.922	1.090
3	Yes	.90 <sup>b</sup>	.338	.506	.605	.719	.782	.836	.981
4	Yes	.90	.324	.501	.611	.743	.819	.886	1.085
2	No	.95	.384	.563	.666	.780	.842	.895	1.035
3	Yes	.95 <sup>b</sup>	.363	.533	.631	.741	.800	.851	.983
4	Yes	.95	.358	.533	.638	.757	.822	.879	1.033

<sup>a</sup> Bias in variance estimates.

<sup>b</sup> Incorrectly assume all correctly identified.

TABLE 4. Apparent weighting factors for group averages in sire evaluation with various combinations of misidentification for intrasire correlation = .08.

Situation	Bias <sup>a</sup>	Fraction correctly identified	Number of cows per group in evaluation						
			10	20	30	50	70	100	1,000
1	No	1.00	.465	.635	.723	.813	.859	.897	.989
2	No	.80	.437	.637	.752	.879	.948	1.006	1.228
3	Yes	.80 <sup>b</sup>	.359	.528	.627	.737	.797	.849	.982
4	Yes	.80	.325	.509	.627	.771	.856	.932	1.214
2	No	.85	.449	.644	.753	.872	.935	.988	1.158
3	Yes	.85 <sup>b</sup>	.383	.554	.651	.757	.813	.861	.984
4	Yes	.85	.361	.549	.663	.797	.872	.938	1.150
2	No	.90	.457	.645	.748	.857	.914	.963	1.095
3	Yes	.90 <sup>b</sup>	.409	.581	.675	.776	.829	.874	.986
4	Yes	.90	.398	.584	.692	.812	.877	.933	1.091
2	No	.95	.462	.642	.737	.837	.888	.931	1.039
3	Yes	.95 <sup>b</sup>	.437	.608	.699	.795	.845	.886	.987
4	Yes	.95	.433	.613	.712	.817	.872	.919	1.038

<sup>a</sup> Bias in variance estimates.

<sup>b</sup> Incorrectly assume all correctly identified.

TABLE 5. Expected values of apparent variances of sire averages (standardized, Model 1).

Situation	Bias in variance estimates	Fraction misidentified in evaluation	Expected value of apparent variance of sire average
1	No	None	$t + (1-t)/n_s$
2	No	Known	$xt + (1-t)/n_s$
3	Yes	Assume none incorrectly	$[x - (1-x)/(n_v-1)]t + \{[n_v(1-x)/(n_v-1)]t + 1-t\}/n_s$
4	Yes	Known	$x[x - (1-x)/(n_v-1)]t + \{[n_v(1-x)/(n_v-1)]t + 1-t\}/n_s$

$t = \sigma_s^2/(\sigma_s^2 + \sigma_e^2)$  = true intrasire correlation coefficient.

$n_s$  = number of records in group average for evaluating a sire.

$n_v$  = number of records in group average for estimating  $\sigma_s^2$  and  $\sigma_e^2$ .

$x = 2p^2 + p - 1$  where  $p$  is the fraction of cows with records in the group average actually sired by that sire.

truncated or censored normal distribution. If biased regression coefficients and variance components are used instead of true  $b$  and  $\sigma_y$  in estimating genetic progress, such estimates can be described as apparent estimates. Similarly the correlation between true and predicted genetic values of sires is usually estimated as  $(b)^{1/2}$  where  $b$  is the correct regression coefficient when all records are correctly identified. If a biased regression coefficient is substituted, the resulting value will be the apparent correlation.

The apparent expected genetic progress as well as apparent correlations between true and predicted genetic values differ for the different

situations. The apparent correlations are the square roots of the regression coefficients in Tables 1, 2, 3, and 4. The apparent standardized variances of the sire average are in Table 5. The equation for Situation 2 gives the unbiased estimate of genetic progress when some records are misidentified. The apparent variance of the sire average for Situation 3 is determined by the estimates of  $\sigma_s^2$  and  $\sigma_e^2$  found from misidentified data. The formula in Table 5 results from substitution of the expected values of the estimates for  $\sigma_s^2$  and  $\sigma_e^2$ . This is the apparent variance of the average in terms of expected values and corresponds to the practical situation. The fourth alterna-

TABLE 6. Apparent estimated genetic progress with various combinations of misidentification for intrasire correlation = .04 relative to no misidentification. Selection intensity factor is assumed to be unity as is the phenotypic standard deviation.

Situation	Bias <sup>b</sup>	Fraction correctly identified	Number of cows per group in evaluation						
			10	20	30	50	70	100	1,000
1 <sup>a</sup>	No	1.00	.108	.135	.149	.164	.173	.180	.198
2 <sup>d</sup>	No	.80	84%	87%	89%	91%	91%	93%	99%
3 <sup>d</sup>	Yes	.80 <sup>c</sup>	69	72	73	76	76	77	79
4 <sup>d</sup>	Yes	.80	57	61	63	67	69	71	78
2 <sup>d</sup>	No	.85	89	90	92	94	94	95	99
3 <sup>d</sup>	Yes	.85 <sup>c</sup>	76	78	79	81	81	82	84
4 <sup>d</sup>	Yes	.85	67	70	72	75	76	78	83
2 <sup>d</sup>	No	.90	93	94	95	96	97	97	99
3 <sup>d</sup>	Yes	.90 <sup>c</sup>	83	84	85	87	87	88	89
4 <sup>d</sup>	Yes	.90	77	79	81	83	84	85	83
2 <sup>d</sup>	No	.95	97	97	98	98	99	99	99
3 <sup>d</sup>	Yes	.95 <sup>c</sup>	92	92	93	93	93	94	94
4 <sup>d</sup>	Yes	.95	88	89	90	91	92	92	94

<sup>a</sup> Expected standardized genetic progress for Situation 1 (no misidentification).

<sup>b</sup> Bias in variance estimates.

<sup>c</sup> Incorrectly assume all correctly identified.

<sup>d</sup> Apparent genetic progress as a percentage of progress for Situation 1.

TABLE 7. Apparent estimated genetic progress with various combinations of misidentification for intrasire correlation = .06 relative to no misidentification. Selection intensity factor is assumed to be unity as is the phenotypic standard deviation.

Situation	Bias <sup>b</sup>	Fraction correctly identified	Number of cows per group in evaluation						
			10	20	30	50	70	100	1,000
1 <sup>a</sup>	No	1.00	.153	.183	.199	.214	.221	.228	.243
2 <sup>d</sup>	No	.80	86%	89%	90%	92%	93%	94%	100%
3 <sup>d</sup>	Yes	.80 <sup>c</sup>	71	73	74	77	78	79	79
4 <sup>d</sup>	Yes	.80	59	63	66	66	71	73	79
2 <sup>d</sup>	No	.85	90	92	93	94	96	96	100
3 <sup>d</sup>	Yes	.85 <sup>c</sup>	77	79	80	82	83	83	84
4 <sup>d</sup>	Yes	.85	68	72	74	77	78	79	84
2 <sup>d</sup>	No	.90	93	95	95	97	98	98	100
3 <sup>d</sup>	Yes	.90 <sup>c</sup>	84	86	86	87	88	89	90
4 <sup>d</sup>	Yes	.90	78	81	82	84	86	86	89
2 <sup>d</sup>	No	.95	97	98	98	99	99	99	100
3 <sup>d</sup>	Yes	.95 <sup>c</sup>	92	93	93	93	94	94	95
4 <sup>d</sup>	Yes	.95	88	90	91	92	93	93	95

<sup>a</sup> Expected standardized genetic progress for Situation 1 (no misidentification).

<sup>b</sup> Bias in variance estimates.

<sup>c</sup> Incorrectly assume all correctly identified.

<sup>d</sup> Apparent genetic progress as a percentage of progress for Situation 1.

TABLE 8. Apparent estimated genetic progress with various combinations of misidentification for intrasire correlation = .08 relative to no misidentification. Selection intensity factor is assumed to be unity as is the phenotypic standard deviation.

Situation	Bias <sup>b</sup>	Fraction correctly identified	Number of cows per group in evaluation						
			10	20	30	50	70	100	1,000
1 <sup>a</sup>	No	1.00	.193	.225	.240	.255	.262	.268	.281
2 <sup>d</sup>	No	.80	87%	90%	91%	93%	94%	95%	100%
3 <sup>d</sup>	Yes	.80 <sup>c</sup>	71	74	76	78	79	79	79
4 <sup>d</sup>	Yes	.80	61	65	68	71	73	74	79
2 <sup>d</sup>	No	.85	91	93	94	95	96	97	100
3 <sup>d</sup>	Yes	.85 <sup>c</sup>	78	80	81	82	83	84	84
4 <sup>d</sup>	Yes	.85	69	73	76	78	79	81	84
2 <sup>d</sup>	No	.90	94	96	97	97	98	98	100
3 <sup>d</sup>	Yes	.90 <sup>c</sup>	84	86	87	88	89	89	90
4 <sup>d</sup>	Yes	.90	79	82	84	85	86	87	90
2 <sup>d</sup>	No	.95	97	98	99	99	99	99	100
3 <sup>d</sup>	Yes	.95 <sup>c</sup>	92	93	93	94	94	94	95
4 <sup>d</sup>	Yes	.95	89	91	92	93	93	94	95

<sup>a</sup> Expected standardized genetic progress for Situation 1 (no misidentification).

<sup>b</sup> Bias in variance estimates.

<sup>c</sup> Incorrectly assume all correctly identified.

<sup>d</sup> Apparent genetic progress as a percentage of progress for Situation 1.

tive is the substitution of the biased estimates into the formula for Situation 2. Situation 4 is the most confusing but may happen if adjustments are made in sire evaluation for proportional misidentification with variance component estimates from misidentified data without considering the misidentification in variance component estimation. Such a procedure should not be followed to compensate for misidentified records in evaluation.

The results of these 4 procedures to estimate genetic progress are in Tables 6, 7, and 8 for true intraclass correlations of .04, .06, and .08. The estimates are a percentage of the estimated progress for complete identification in both estimation of variance components and evaluation of sires. The results are similar for all values of the intrasire correlation. None of the estimates of genetic progress based on the 3 combinations of misidentification is as great as the genetic progress expected with perfect identification. The Situation 2 estimates give the unbiased expected progress when records

are misidentified since those estimates are based on the parameter values of  $\sigma_s^2$  and  $\sigma_e^2$  and are greater than either Situation 3 or Situation 4 estimates. Situation 3 and Situation 4 estimates are similar especially as the fraction misidentified becomes smaller.

### Conclusions

Biases in evaluation of sires will result from misidentified records in estimating  $\sigma_s^2$  and  $\sigma_e^2$ . The biases are greater for larger fractions misidentified than for smaller fractions misidentified.

Genetic progress will be underestimated by variance components estimated from misidentified records. The extent of the underestimation depends primarily on the fraction of records which are misidentified.

### Reference

- (1) Van Vleck, L. D. 1970. Misidentification in estimating the paternal sib correlation. *J. Dairy Sci.*, 53: 1469.