EFFECT OF MISIDENTIFICATION ON THE ESTIMATION OF BREEDING VALUE AND HERITABILITY IN CATTLE

H. Geldermann, U. Pieper and W. E. Weber
Institut für Tierzucht und Vererbungsforschung
der Tierärztlchen Hochschule Hannover
Hannover, Federal Republic of Germany

ABSTRACT

For daughter groups of 15 test bulls, controls of paternity were performed by using blood group factors and biochemical polymorphisms. Data of incorrectly assigned daughters influenced the estimation of breeding values, heritabilities and correlations for milk performance traits. Formulae are given that show the effects of variable misidentification rates on estimation of breeding values, selection intensities, heritabilities, and genetic gains. For example, for milk fat yield, the genetic gains drop at a misidentification rate of 15% between 8.7% (for $h^2 = .5$) and 16.9% (for $h^2 = .2$) below values attained without misidentifications. Consequently, decreasing misidentification rates in progeny of test bulls can be used 1) to diminish the progeny size per test bull for constant genetic gain, 2) to achieve more precise ranking of all or distinct test bulls according to their "true" breeding values and(or) 3) to increase the number of test bulls by using the same amount of test inseminations and the same precision of ranking. Actions to reduce misidentification rates in cattle populations are discussed.

(Key Words: Cattle, Breeding Value, Heritability, Selection, Genetic Gain, Identification.)

Introduction

By using parental control procedures, incorrectly designated parents and offspring can be recognized in breeding populations. In several investigations on cattle (Koch et al., 1970; Lütich, 1973; Erhard, 1979; Geldermann et al., 1983), between 4 and 23% of the paternal identifications were found to be false. Therefore, estimation of breeding values and heritabilities on the basis of progeny data also include individuals of questionable paternity. This can impair the accuracy and precision of the estimated parameters. The extent to which this occurred is given for daughter groups of 15 test bulls.

Materials and Methods

From November 1981 to April 1982 blood samples were randomly collected from 1,221 German Friesian heifers that were registered as daughters of 15 test bulls of the insemination period 1977 to 1978. The cattle were located in milk performance controlled farms in Lower Saxony (Federal Republic of Germany). Blood samples were also taken from the registered sires and were tested for the blood group factors of the systems A, B, C, F, J, L, M, S, Z and R'. (Dr. J. Koch, Abt. Blutgruppen des Tierärztlichen Instituts der Universität Göttlingen). In addition, electrophoretic analyses (Thinnes et al., 1976; Gebicke and Geldermann, 1977) were used to reveal the types of postalbumines, transferrins, posttransferrins and amylases-1. Under the given conditions for parental controls an exclusion rate of about 78% (Kaup, 1983) was attained.

Data of milk performance for the first lactations were prepared by the computer center responsible for evaluation of breeding values. Breeding values were also calculated for the sires (Lederer and Buthmann, 1982) using all registered "daughters", or only those daughters that could not be excluded after parental control. Estimation of heritabilities
and correlations were carried out by the LSML-76 program of Harvey (1978) in the paternal half-sib groups.

Results and Discussion

The data of first lactation performances of the 15 daughter groups are summarized in table 1. In total, the incorrectly assigned animals that were detected produced less milk than the others, but had higher milk fat content. Such a comparison revealed no clear results for single sires, because the number of daughters per sire was small. Only in a few cases were the values of first lactations of the incorrectly designated cows significantly different from the non-excludable daughters.

Table 2 shows breeding values of test bulls estimated using all the investigated daughters (a) or only the non-excludable daughters (b). Both approaches resulted in remarkable differences among some bulls, especially sire no. 5. The differences between breeding values estimated with or without the incorrectly identified individuals, influenced the rank order of the test bulls. For example, according to milk fat yield, sire no. 5 ranks 11th if all the investigated daughters are considered but fourth if only the non-excludable daughters are used. On the average, breeding values derived from all investigated daughters differed little from those of non-excludable daughters. For that reason, both evaluations of the breeding values correlate closely for all three milk performance traits (table 2). However, breeding values for milk fat content and milk fat yield tend to vary more if using only non-excludable daughters compared with the whole groups of cows investigated.

Data from daughters of the 15 test bulls led to the estimates of heritabilities and correlations given in table 3. Compared with estimates computed from all investigated individuals, estimates from the non-excludable daughters were larger for heritabilities and, in some cases, for correlations.

The results show that misidentifications in daughter groups of test bulls influence the estimates of breeding parameters. In the investigated daughter groups a misidentification rate of 13.2% was found (table 1). If the exclusion rate is 78% for paternal pedigree errors using polymorphic characters and sire-offspring pairs (Kaup, 1983), an overall error rate of 16.7% (13.2/.78) is expected, so that 3.5% of misidentifications remain undetected. Thus, the differences might have been greater if all misidentifications could have been discovered.

For breeding purposes it is important to know how pedigree errors influence the precision of progeny tests and diminish the selection response. Not only misidentification of daughters to sires, but also of daughters to dams, can produce errors in estimating breeding values and heritabilities. As described by Van Vleck (1970a) the expected covariance between daughters and dams equals \((1 - p) \text{Cov}(\text{daughter, dam})\), if \(p\) is the frequency of misidentification and \(\text{Cov}(\text{daughter, dam})\) is the true covariance between daughters and dams. But misidentifications between daughters and dams are much less frequent than those between daughters and sires (Koch et al., 1970) so that in the following only the paternal pedigree errors are considered.

Precision of Estimating Breeding Values by Progeny Tests. In dairy cattle breeding, progeny tests are of central importance. The precision of these tests depends on number of offspring per sire, heritability of the trait, and the fraction of incorrectly identified individuals. Van Vleck (1970b) and Christensen et al. (1982) pointed out that the precision of estimated breeding value on the basis of phenotypic values of progeny can be measured by the coefficient of determination.

\[
\hat{\sigma}_{IA}^2 = \frac{(1 - p)^2 n h^2/4}{1 + (1-p)((1-p)n - 1) h^2/4} \quad [1]
\]

with

\( p = \text{frequency of incorrectly identified offspring}, \)
\n\( h^2 = \text{heritability and} \)
\n\( n = \text{group size}. \)

As seen, the coefficient of determination depends on the frequency of misidentification. For a given value of \(p\), \(\hat{\sigma}_{IA}^2\) increases with increasing values for heritability and group size (Christensen et al., 1982). For example, in case of \(p = .15, h^2 = .25\) and \(n = 80\), the coefficient of determination is .92 in accordance with results in table 2.

The value of \(\hat{\sigma}_{IA}^2\) yields no measurement for the deviation between estimated and true breeding value for a single test bull. But, this deviation can influence the rank order of test bulls according to their breeding values, as
### Table 1. Milk Performance of First Lactation

<table>
<thead>
<tr>
<th>Sire No.</th>
<th>Line&lt;sup&gt;a&lt;/sup&gt;</th>
<th>n</th>
<th>Milk yield (kg)</th>
<th>Fat content (%)</th>
<th>Fat yield (kg)</th>
<th>Protein yield (%)</th>
<th>Protein yield (kg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>a</td>
<td>82</td>
<td>4,925</td>
<td>1,005</td>
<td>3.73</td>
<td>0.31</td>
<td>183.8</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>4</td>
<td>5,360</td>
<td>1,056</td>
<td>3.57</td>
<td>0.59</td>
<td>188.6</td>
</tr>
<tr>
<td>2</td>
<td>a</td>
<td>72</td>
<td>5,198</td>
<td>981</td>
<td>3.84</td>
<td>0.37</td>
<td>197.3</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>11</td>
<td>4,963</td>
<td>730</td>
<td>3.96</td>
<td>0.53</td>
<td>198.3</td>
</tr>
<tr>
<td>3</td>
<td>a</td>
<td>67</td>
<td>4,942</td>
<td>973</td>
<td>3.89</td>
<td>0.32</td>
<td>191.5</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>5</td>
<td>5,215</td>
<td>1,127</td>
<td>3.88</td>
<td>0.41</td>
<td>201.2</td>
</tr>
<tr>
<td>4</td>
<td>a</td>
<td>44</td>
<td>4,605</td>
<td>1,093</td>
<td>3.70</td>
<td>0.29</td>
<td>171.1</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>14</td>
<td>4,613</td>
<td>1,755</td>
<td>3.92</td>
<td>0.41&lt;sup&gt;b&lt;/sup&gt;</td>
<td>180.6</td>
</tr>
<tr>
<td>5</td>
<td>a</td>
<td>68</td>
<td>5,142</td>
<td>1,036</td>
<td>3.93</td>
<td>0.34</td>
<td>201.7</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>12</td>
<td>4,665</td>
<td>846</td>
<td>4.10</td>
<td>0.55</td>
<td>190.4</td>
</tr>
<tr>
<td>6</td>
<td>a</td>
<td>60</td>
<td>4,941</td>
<td>1,036</td>
<td>3.92</td>
<td>0.38</td>
<td>194.7</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>15</td>
<td>4,675</td>
<td>1,367</td>
<td>3.89</td>
<td>0.35</td>
<td>182.6</td>
</tr>
<tr>
<td>7</td>
<td>a</td>
<td>47</td>
<td>4,853</td>
<td>866</td>
<td>3.77</td>
<td>0.28</td>
<td>183.3</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>6</td>
<td>4,898</td>
<td>673</td>
<td>4.01</td>
<td>0.26&lt;sup&gt;b&lt;/sup&gt;</td>
<td>196.5</td>
</tr>
<tr>
<td>8</td>
<td>a</td>
<td>92</td>
<td>4,967</td>
<td>1,377</td>
<td>3.90</td>
<td>0.39</td>
<td>194.4</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>7</td>
<td>4,956</td>
<td>768</td>
<td>4.11</td>
<td>0.29</td>
<td>204.3</td>
</tr>
<tr>
<td>9</td>
<td>a</td>
<td>57</td>
<td>4,826</td>
<td>929</td>
<td>3.65</td>
<td>0.41</td>
<td>177.6</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>9</td>
<td>4,527</td>
<td>1,403</td>
<td>4.06</td>
<td>0.30&lt;sup&gt;c&lt;/sup&gt;</td>
<td>183.9</td>
</tr>
<tr>
<td>10</td>
<td>a</td>
<td>80</td>
<td>5,285</td>
<td>1,141</td>
<td>3.93</td>
<td>0.37</td>
<td>206.6</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>12</td>
<td>4,244</td>
<td>1,510</td>
<td>3.90</td>
<td>0.43</td>
<td>188.2</td>
</tr>
<tr>
<td>11</td>
<td>a</td>
<td>71</td>
<td>5,872</td>
<td>1,148</td>
<td>3.67</td>
<td>0.48</td>
<td>215.5</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>20</td>
<td>5,557</td>
<td>1,363</td>
<td>3.79</td>
<td>0.48</td>
<td>208.5</td>
</tr>
<tr>
<td>12</td>
<td>a</td>
<td>87</td>
<td>5,460</td>
<td>1,078</td>
<td>3.93</td>
<td>0.38</td>
<td>215.3</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>11</td>
<td>5,168</td>
<td>711</td>
<td>4.17</td>
<td>0.40&lt;sup&gt;b&lt;/sup&gt;</td>
<td>215.6</td>
</tr>
<tr>
<td>13</td>
<td>a</td>
<td>81</td>
<td>5,065</td>
<td>1,212</td>
<td>4.09</td>
<td>0.43</td>
<td>209.2</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>7</td>
<td>4,462</td>
<td>727</td>
<td>4.18</td>
<td>0.51</td>
<td>186.6</td>
</tr>
<tr>
<td>14</td>
<td>a</td>
<td>75</td>
<td>5,005</td>
<td>1,105</td>
<td>3.82</td>
<td>0.30</td>
<td>191.2</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>15</td>
<td>4,694</td>
<td>1,216</td>
<td>3.92</td>
<td>0.40</td>
<td>184.1</td>
</tr>
<tr>
<td>15</td>
<td>a</td>
<td>79</td>
<td>5,013</td>
<td>999</td>
<td>3.71</td>
<td>0.35</td>
<td>185.6</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>11</td>
<td>5,376</td>
<td>940</td>
<td>3.89</td>
<td>0.35</td>
<td>208.9</td>
</tr>
<tr>
<td>Total</td>
<td>a</td>
<td>1,062</td>
<td>5,099</td>
<td>1,115</td>
<td>3.84</td>
<td>0.38</td>
<td>196.0</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>159</td>
<td>4,933</td>
<td>1,108&lt;sup&gt;b&lt;/sup&gt;</td>
<td>3.95</td>
<td>0.43&lt;sup&gt;d&lt;/sup&gt;</td>
<td>194.5</td>
</tr>
</tbody>
</table>

<sup>a</sup>Values of daughters non-excludable from the sire concerned (a) and values of animals incorrectly identified to the sire concerned (b).

<sup>b</sup>Values in line a different from line b (P<.05).

<sup>c</sup>Values in line a different from line b (P<.01).

<sup>d</sup>Values in line a different from line b (P<.001).
### TABLE 2. ESTIMATED BREEDING VALUES

<table>
<thead>
<tr>
<th>Sire No.</th>
<th>All investigated daughters</th>
<th>Non-excludable daughters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(a)</td>
<td>(b)</td>
</tr>
<tr>
<td></td>
<td>Breeding values</td>
<td>Breeding values</td>
</tr>
<tr>
<td></td>
<td>Milk yield</td>
<td>Fat content</td>
</tr>
<tr>
<td></td>
<td>kg</td>
<td>Ranking</td>
</tr>
<tr>
<td>1</td>
<td>86</td>
<td>470</td>
</tr>
<tr>
<td>2</td>
<td>83</td>
<td>523</td>
</tr>
<tr>
<td>3</td>
<td>72</td>
<td>323</td>
</tr>
<tr>
<td>4</td>
<td>58</td>
<td>239</td>
</tr>
<tr>
<td>5</td>
<td>80</td>
<td>335</td>
</tr>
<tr>
<td>6</td>
<td>75</td>
<td>361</td>
</tr>
<tr>
<td>7</td>
<td>53</td>
<td>96</td>
</tr>
<tr>
<td>8</td>
<td>99</td>
<td>494</td>
</tr>
<tr>
<td>9</td>
<td>66</td>
<td>144</td>
</tr>
<tr>
<td>10</td>
<td>92</td>
<td>361</td>
</tr>
<tr>
<td>11</td>
<td>91</td>
<td>1,071</td>
</tr>
<tr>
<td>12</td>
<td>98</td>
<td>463</td>
</tr>
<tr>
<td>13</td>
<td>88</td>
<td>52</td>
</tr>
<tr>
<td>14</td>
<td>90</td>
<td>287</td>
</tr>
<tr>
<td>15</td>
<td>90</td>
<td>334</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sire No.</th>
<th>Sum</th>
<th>X</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,221</td>
<td></td>
<td>81.4</td>
<td>14.0</td>
</tr>
<tr>
<td></td>
<td>1,062</td>
<td>368</td>
<td>241</td>
</tr>
</tbody>
</table>

Coefficient of determination ($r^2$), (a) with (b): .983, .904, .848
TABLE 3. HERITABILITIES AND CORRELATIONS

<table>
<thead>
<tr>
<th>Trait</th>
<th>Line</th>
<th>Milk yield</th>
<th>Fat content</th>
<th>Fat yield</th>
<th>Protein content</th>
<th>Protein yield</th>
</tr>
</thead>
<tbody>
<tr>
<td>Milk yield</td>
<td>a</td>
<td>.212</td>
<td>-.023</td>
<td>.901</td>
<td>.012</td>
<td>.887</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>.218</td>
<td>.001</td>
<td>.906</td>
<td>.019</td>
<td>.886</td>
</tr>
<tr>
<td>Fat content</td>
<td>a</td>
<td>-.034</td>
<td>.320</td>
<td>.402</td>
<td>.242</td>
<td>.096</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>.051</td>
<td>.381</td>
<td>.413</td>
<td>.229</td>
<td>.114</td>
</tr>
<tr>
<td>Fat yield</td>
<td>a</td>
<td>.839</td>
<td>.514</td>
<td>.218</td>
<td>.111</td>
<td>.851</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>.841</td>
<td>.581</td>
<td>.247</td>
<td>.109</td>
<td>.851</td>
</tr>
<tr>
<td>Protein content</td>
<td>a</td>
<td>-.101</td>
<td>.573</td>
<td>.250</td>
<td>.175</td>
<td>.424</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>-.106</td>
<td>.487</td>
<td>.209</td>
<td>.214</td>
<td>.434</td>
</tr>
<tr>
<td>Protein yield</td>
<td>a</td>
<td>.907</td>
<td>.204</td>
<td>.896</td>
<td>.315</td>
<td>.144</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>.893</td>
<td>.288</td>
<td>.893</td>
<td>.343</td>
<td>.154</td>
</tr>
</tbody>
</table>

a) Diagonal: heritabilities; above the diagonal: phenotypic correlations; below the diagonal: genotypic correlations.

b) Values including all investigated daughters (a) and values including only non-excludable daughters (b).

shown in table 2. Gundel and Geldermann (1980) examined the effect of misidentification rates on the probability of ranking errors for progeny tested sires. They found that with a misidentification rate of 15% the same ranking precision can be attained with about 25% smaller group size, if the pedigree errors are detected and removed from the progeny groups.

Misidentification is expected to reduce the differences between the estimated breeding values, i.e., the estimates of breeding values of inferior sires are increased, whereas those of superior sires are reduced. This is supported by the data of table 2 for the breeding values of milk fat content and milk fat yield. The precision of estimating sires with extreme breeding values is impaired particularly. This holds true especially for superior sires if the average performance of the incorrectly assigned offspring is less than the population mean (table 1). In addition, when misidentifications are unequally distributed among the progeny groups (table 1), the more the distribution varies, the more the rank order is affected.

Given a specific number of sires in a testing program, the misidentification rate has an effect on the probability of selecting sires with superior "true" breeding values. Even without misidentifications, the breeding value of a sire can only be estimated. Besides non-additive and environmental effects, the unknown breeding values of dams also are included in the estimation. Therefore, assuming that for sire i there are produced n offspring by random mating with n dams, the estimated breeding value is:

\[ \hat{A}_i = 2 \cdot \overline{\text{progeny}} \]

\[ = A_{ti} + \overline{A_{mi}} + E_i, \]

with

\[ A_{ti} = \text{the "true" breeding value of sire i, with} \]

\[ E(A_{ti}) = \mu_t > 0, \text{because the sires have} \]

\[ \overline{A_{mi}} = \text{mean of the breeding values of dams,} \]

\[ \text{with } E(\overline{A_{mi}}) = 0 \text{ and} \]

\[ E_i = \text{mean of other effects, with } E(E_i) = 0. \]

All effects are assumed to be normally distributed.

The variances are:

\[ \text{Var} (A_{ti}) = \sigma^2, \]

\[ \text{Var} (\overline{A_{mi}}) = \frac{1}{n} \sigma^2, \text{if the mean of n breeding} \]

\[ \text{values is regarded and} \]

\[ \text{Var} (E_i) = \frac{4\sigma^2}{n}, \text{because the breeding values} \]

\[ \text{are obtained by doubling the mean} \]

\[ \text{of progeny.} \]

From h^2 = \frac{\sigma_A^2}{\sigma_A^2 + \sigma_E^2} \text{ follows } \frac{4\sigma_E^2}{\sigma_A^2 + \sigma_E^2} = \frac{4(1 - h^2)}{nh^2} \sigma_A^2.

In case of a misidentification rate p the portion of the true breeding value is changed to (1-p)A_{ti} + p\overline{A_{mi}}, with \overline{A_{mi}} as mean breeding value of all unknown sires of the incorrectly identified progeny. With p x n progeny from p x n unknown sires we get:
So, if misidentifications are present, the breeding value is estimated as
\[
\hat{A}_i = (1-p) A_{hi} + pA_{fi} + A_m + E_i, \quad [2]
\]
with the variance
\[
\text{Var} (\hat{A}_i) = \left[ (1-p)^2 + \frac{p+1}{n} \right] \sigma_A^2 + \frac{4\sigma_E^2}{n} = \left[ (1-p)^2 + \frac{p+1}{n} + \frac{4(1-h^2)}{n^2 h^2} \right] \sigma_A^2.
\]

since \( A_{hi}, A_{fi}, A_m \) and \( E_i \) are assumed to be uncorrelated.

In the following, we consider how the number \( m \) of sires to be selected depends on the misidentification rate \( p \) under the condition that the best \( k \leq m \) sires with highest breeding values are included with given probability
\[
P > 1 - \alpha = 1 - (\alpha_1 + k\alpha_2).
\]
The critical value of the selection is the sum of two \( p \)-dependent values, \( u \) and \( w \) (see appendix), so that an individual \( i \) is selected if it has a breeding value \( \hat{A}_i \geq u + w \). The \( k \)-highest “true” breeding value multiplied by \( 1-p \) is less than \( u \) with a probability \( \alpha_1 \); for sire \( i \) the term \( pA_{fi} + A_m + E_i \) is less than \( w \) with a probability \( \alpha_2 \).

Considering \( N \) test sires with \( n \) daughters per sire, for large \( N \) we get
\[
\frac{m}{N} \approx \Phi \left( \frac{u}{\sigma_t} = P(\hat{A}_i > u + w) = 1 - \Phi \left( \frac{x}{\sigma_x} \right) = \Phi \left( \frac{x}{\sigma_x} \right) \right)
\]
with
\[
\phi = \text{function of a standardized normal distribution},
\]
\[
x = u + w - (1-p) \mu_t
\]
and
\[
\sigma_x^2 = \text{Var} (\hat{A}_i).
\]

By defining
\[
\bar{u} = \frac{1-p}{\sigma_t} \mu_t \quad \text{and} \quad \bar{w} = \frac{w}{\sigma_w},
\]
and using
\[
\sigma_x^2 = \sigma_A^2 + \sigma_w^2 = \frac{p+1}{n} \sigma_A^2 + \frac{4\sigma_E^2}{n} \quad \text{and} \quad \sigma_x^2 = \sigma_A^2 \frac{p+1}{n} + \frac{4\sigma_E^2}{n^2 h^2},
\]
it follows that
\[
\frac{m}{N} \approx \phi \left( \frac{(1-p)\bar{u} + \frac{p+1}{n} + \frac{4\sigma_E^2}{n^2 \sigma_A^2} \bar{w}}{\sqrt{(1-p)^2 + \frac{p+1}{n} + \frac{4\sigma_E^2}{n^2 \sigma_A^2}}} \right).
\]

If \( N, n, k, p, \mu_t \) and \( h^2 \) are known, the values for \( \bar{u} \) and \( u \) at given \( \alpha_1 \) and the values for \( \bar{w} \) and \( w \) at given \( \alpha_2 \) can be calculated (see appendix). Then \( \frac{m}{N} \) can be computed with equation [3]. The selection intensity \( i \) depends on \( \frac{m}{N} \) and is smaller for larger values of \( m \).

**Precision of Estimating Heritabilities.** Incorrectly identified offspring influence the estimated heritability values. If we assume 1) equal progeny sizes with equal misidentification rates per sire, 2) incorrectly identified individuals as offspring from many other sires and therefore as a random sample from the population and 3) equal phenotypic variance for correctly and incorrectly assigned offspring, the expectation of heritability is
\[
\hat{h}^2 = \frac{n(1-p)^2 - 1 + p}{n-1} \cdot h^2
\]
(Van Vleck, 1970a), [4]
with
\[ n = \text{group size}, \]
\[ p = \text{misidentification rate and} \]
\[ h^2 = \text{"true" heritability}. \]

Figure 1 shows the dependence of estimated heritabilities on misidentification rates. As defined previously, the calculations assumed equal misidentification rates for all progenies, whereas different rates of misidentifications per progeny diminish the bias on the estimated heritabilities (Van Vleck, 1970a). This explains why in our experimental results the estimated values of heritability decrease only up to 22% if the material contains, instead of 3.5%, an error rate of 16.7% (see table 3).

**Effect of Misidentification on Genetic Gain.**

The genetic gain in dairy cattle is based on the four paths: sire-son, sire-daughter, dam-son and dam-daughter (Rendel and Robertson, 1950). The improvement for each path is given by

\[ \Delta G = r_{IA} \cdot i \cdot \sigma_A, \]

with
\[ r_{IA} = \text{correlation between the breeding value and the phenotypic mean of the progeny}, \]
\[ i = \text{intensity of selection and} \]
\[ \sigma_A = \text{additive genetic standard deviation}. \]

In this paper, the sire-daughter path is evaluated. With a rate p of misidentification, \( r_{IA} \) (see equation 1) and i are reduced. The reduction in i is caused by a larger value of N (see equation 3). For genetic gain of milk fat yield under the conditions of N = 300, k = 10, n = 80 and \( \sigma_A = 40 \text{ kg} \), the relation between \( \Delta G \) and p can be calculated for different \( h^2 \). As shown in figure 2, at a misidentification rate of 15%, the genetic gain drops between 8.7% (for \( h^2 = .5 \)) and 16.9% (for \( h^2 = .2 \)) below the values attained without misidentification.

For practical breeding purposes, one may ask to what degree the number of offspring can be reduced, if no misidentifications occur (by parental controls), so that \( \Delta G \) should remain constant. For the previous example, that relation is given in figure 3. Different values of heritability have only minor influences.

---

**Figure 1.** Estimation of heritabilities in case of variable misidentification rates and a group size of 80 offspring per sire. The hatched lines connect the values found from the investigations for milk fat content (○), milk fat yield (●), milk yield (♦), milk protein content (★) and milk protein yield (★).

**Figure 2.** Development of the genetic gain for variable misidentification rates and the assumptions stated in the text.
Conclusions

In cattle breeding, misidentification influences estimates of heritability and breeding values, when progeny groups are used. Misidentification reduces genetic gain and (or) increases the size of progeny groups of test bulls necessary for a given gain. Consequences of a decrease in misidentification rates in progenies of test bulls are: 1) diminishing the progeny size per test bull for a constant genetic gain, 2) more precise ranking of all or distinct test bulls according to their "true" breeding values and (or) 3) increasing the number of test bulls by using the same number of test inseminations and the same precision of ranking.

Achieving these goals will produce an economical advantage. As described by Kaup (1983) and Tiessen (1983), the breeders' interest and the care of the inseminating personnel influence strongly the number of misidentifications. Consequently, in animals, for which the owners do not expect parental controls, higher misidentification rates occur than in others (Geldermann et al., 1983).

Depending on the causes, misidentification rates can be reduced by appropriate actions. First, steps should be taken that help to avoid errors. For example, mini-computers can be used for registration and transmission of insemination and calving data. In addition, multiple inseminations with different bulls should be avoided. Most importantly, special efforts should be made to train the personnel concerned with insemination and identification. Furthermore, the practical breeders should be convinced of the necessity of correct data recording.

Finally, misidentifications can be recognized with parental controls and removed from the material. Three methods can be employed for checking on progeny groups of test bulls:

First, paternity controls of all individuals whose performances should be used for estimation of breeding values can be effected by including offspring of all test bulls. In this case, low exclusion rates are sufficient (Gundel and Geldermann, 1980). So for economic reasons only a few gene systems, selected separately for each sire, should be used. For saving costs when milk samples are taken, aliquots from routine milk tests can be investigated in special laboratories.

Second, paternity controls can be used in progenies of selected test bulls using all individuals whose performances should be used for estimation of breeding values. For this, the sires included should have the highest estimated breeding values on the basis of all registered offspring. That action would include sires whose estimated breeding values are especially influenced by misidentifications. Moreover, these sires should be used later on in the population via artificial insemination so that precise estimates are particularly important.

Third, paternity controls can employ a sample of offspring whose performances should be used for estimation of breeding values. Thereby, the individuals can be chosen so that factors which influence the misidentification rates (e.g., inseminating staff, multiple inseminations with different bulls, farms for fattening the bulls) can be checked. The success of these actions depends on improving the accuracy of the pedigree data in the whole breeding population by training. If that can be ensured only a small sample survey will reduce the error rate sufficiently. Sample surveys however, have some disadvantages, i.e., the time-lag between detection and correction of misidentifications, the difficulty of measuring the reduction and a final error rate that probably will be higher than with the other two methods.
Moreover, in practice the sampling can lead to a misrepresentation of the population, impairing the results and escaping immediate detection. This turned out to be the case in empirical data studied by Kaup (1983).

Appendix

A group of sires is selected on the basis of the estimated breeding values \( A_i \). This group should include the \( k \) individuals that have the highest "true" breeding values with a probability \( > 1 - \alpha \). For calculating \( \alpha \), the quantities \( \alpha_1 \) and \( \alpha_2 \) greater than 0 are introduced so that

\[
\alpha = \alpha_1 + k \cdot \alpha_2.
\]

All sires that have a breeding value of

\[
\hat{A}_i = u + w
\]

are placed in this group.

In the following, the determination of \( u \) and \( w \) are explained.

\[
\hat{A}_i = (1 - p) \cdot A_{ti} + p \bar{A}_{fi} + \bar{A}_{mi} + \bar{E}_i
\]

is given in equation [2].

If \( A_{t[1]} \) represent the rank values of the "true" breeding values so that for \( N \) sires,

\[
A_{t[1]} < A_{t[2]} < \ldots < A_{t[N]},
\]

then for \( \alpha_1 \) a value \( u \) can be defined, so that with misidentification rate, \( p \), the probability is,

\[
P[(1 - p) A_{t[N - k + 1]} < u] < \alpha_1.
\]

After the transformation \( \bar{u} = \frac{u}{1 - p} - \mu_t \) this is equivalent to

\[
\frac{N - k + 1}{k} \cdot \frac{1 - \phi(\bar{u})}{\phi(\bar{u})} = F_{\alpha_1} \left[2k, 2(N - k + 1)\right]
\]

(Hald, 1960),

with \( \phi = \) function of a standardized normal distribution and \( F \left[2k, 2(N - k + 1)\right] = F \)-distribution with 2 \( k \) and 2\((N - k + 1)\) degrees of freedom. Then from

\[
\frac{1}{1 + \frac{k}{N - k + 1} \cdot F_{\alpha_1} \left[2k, 2(N - k + 1)\right]},
\]

\( u \) can be determined by

\[
u = (\bar{u} \sigma_t + \mu_t) \cdot (1 - p)\]

For the determination of \( w \) we use

\[
P(p \bar{A}_{fi} + \bar{A}_{mi} + \bar{E}_i < w) = \phi(\frac{w}{\sigma_w}) < \alpha_2,
\]

with

\[
\sigma_w^2 = \frac{p + 1}{n} \cdot \sigma_A^2 + \frac{4 \sigma_E^2}{n} + \left[ \frac{p + 1 + 4(1 - h^2)}{nh^2} \right] \sigma_A^2
\]

(see text).

In total, the probability, not to have all \( k \) individuals with highest "true" breeding values in the selected group, is less or equal than

\[
P[(1 - p) A_{t[N - k + 1]} < u] + \sum_{i=1}^{k} P(p \bar{A}_{fi} + \bar{A}_{mi} + \bar{E}_i < w) < \alpha_1 + k \alpha_2
\]

The actual probability for a selection of all \( k \) superior individuals may be significantly larger than \( 1 - (\alpha_1 + k \alpha_2) \).

Literature Cited

Christensen, L. G., P. Madsen and J. Petersen. 1982. The influence of incorrect sire-identification on
the estimates of genetic parameters and breeding values. Proc. 2nd World Cong. on Genetics Applied to Livestock Production 7:200.