ESTIMATION OF ADDITIVE GENETIC VARIANCE WHEN BASE POPULATIONS ARE SELECTED

J.H.J. van der Werf and I.J.M. de Boer

Wageningen Agricultural University, 6700 AH Wageningen, The Netherlands

ABSTRACT

A population of size 40 was simulated 1,000 times for 10 generations. Five out of twenty males were selected each generation, and each male was mated to four females to have two progeny. The additive genetic variance ($\sigma^2_g$) before selection was 10, and the initial heritability was .5. Due to covariances among animals, inbreeding and gametic disequilibrium, the genetic variance was reduced to 6.72 after 10 generations of selection. Reduction of variance was lower in another population simulated with size 400 and 10% of the males selected. Restricted Maximum Likelihood was used to estimate $\sigma^2_g$ using an animal model. The estimate of $\sigma^2_g$ was empirically unbiased when all data and all relationships were used. Omitting data from selected ancestors caused biased estimates of $\sigma^2_g$ due to not accounting for all gametic disequilibrium. Including additional relationships between assumed base animals adjusted for inbreeding and for covariances. Bias from gametic disequilibrium decreased slightly with the use of more relationship information, and it was smaller in the small population and/or when selection had been practiced for just a few generations.

(Key Words: Variance Components, Estimation, Genetic Variance, Selection.)


Introduction

Parameters for production traits often are estimated from data on selected animals. In principle, selection can be accommodated by an appropriate model that includes all data upon which selection decisions were based (Henderson, 1975; Sorensen and Kennedy, 1984b; Gianola and Fernando, 1986). However, the time span covered by data generally is limited and, in practice, all data since the start of selection are not available.

Sorensen and Kennedy (1984b) simulated several generations of selection and omitted data from earlier generations. They concluded that the estimate of the additive genetic variance before selection was nearly unbiased when their model acknowledged all relationships that developed in previous generations. This seems inconsistent with the condition that all data are needed for unbiased estimation.

Because information on pedigree usually does not date back to a true base population, an assumption concerning a base population is made (e.g., the first generation of animals with data are considered to be unrelated, unselected and non-inbred base animals). Sorensen and Kennedy (1984b) indicated that such a model would estimate the genetic variance in the implied base generation. However, they analyzed data from only two generations with an animal model. The estimate would be the equivalent of offspring on parent regression that is unbiased by selection of parents. This may not be true if more generations are included.

The aim of this research was to study estimates of additive genetic variance when base populations were selected. We examine how estimates are affected by omitting data...
from selected ancestors and by ignoring known relationships of selected animals and then systematically including more relationship information among animals.

**Materials and Methods**

**Simulated Data**

This study follows the Monte Carlo simulation strategy described by Sorensen and Kennedy (1984a). They assumed a genetic model in which a large number of unlinked loci contribute to the genetic variance. A number of s males and m females, all assumed unrelated and unselected, were sampled randomly from a conceptually infinite base population. The base animals were mated at random (ds = d females per male) to produce m males and m females, which is generation 1. The s phenotypically best males were selected in each of the subsequent generations. Each of the females was mated randomly to one of the selected males. Selection was only on males and generations did not overlap.

The model used for simulation of a record for the i\(^{th}\) animal was 
\[
y_i = \mu + a_i + e_i
\]
where \(y_i\) is the phenotypic value of the \(i^{th}\) animal, \(a_i\) is its additive genetic value, and \(e_i\) is a residual random value for possible nonadditive and environmental effects. The parameter \(\mu\) is the phenotypic mean of generation 1. Random values for \(e_i\) were drawn from a normal distribution with mean zero and variance 10. The additive genetic variance before selection \((\sigma^2_a)\) was 10 and initial heritability was .5. For the base animals (generation 0) genetic values were drawn from N[0, 10]. Genetic values for animals of later generations were simulated as:

\[
a_i = \frac{1}{2}a_s + \frac{1}{2}a_d + \phi_i
\]

where \(a_s\) and \(a_d\) are genetic values of sire and dam of individual \(i\). The value \(\phi_i\) results from Mendelian sampling, which is independent of \(a_s\) and \(a_d\) (Bulmer, 1971). The coefficient of inbreeding as defined by Falconer (1989) for the \(i^{th}\) animal is \(F_i\). The variance of \(a_i\) can be presented as

\[
\text{var}(a_i) = (1 + F_i)\sigma^2_a
\]

The variance of \(\phi\) can be given as

\[
\text{var}(\phi) = \frac{1}{4}\text{var}(a_s) + \frac{1}{4}\text{var}(a_d)
\]

\[
+ \frac{1}{2}\text{cov}(a_s,a_d) + \text{var}(\phi)
\]

\[
= \frac{1}{4}\text{var}(a_s) + \frac{1}{4}\text{var}(a_d)
\]

\[
+ \frac{1}{2}\text{cov}(a_s,a_d)
\]

\[
= (1 + F_i)\sigma^2_a - \frac{1}{4}\text{var}(a_s) - \frac{1}{4}\text{var}(a_d)
\]

\[
- \frac{1}{2}\text{cov}(a_s,a_d)
\]

\[
= [(1 + F_i) - \frac{1}{4}(1 + F_s) - \frac{1}{4}(1 + F_d)] - F_i\sigma^2_a
\]

\[
= \frac{1}{2}[1 - \frac{1}{2}(F_s + F_d)]\sigma^2_a
\]

where \(F_s\) and \(F_d\) are inbreeding coefficients for sire and dam, respectively. Inbreeding coefficients were computed using Quaas’ (1976) algorithm. The residual genetic value for each animal \(\phi\) was drawn from a normal distribution with mean zero and variance according to [3]. The number of replicates per population depended on the size and number of generations \((s mg)\) simulated. Per sample, there was one record available for each of \(2s mg\) animals and the \((s + m)\) unselected and unrelated base animals were identified as parents without records (generation 0).

**Analyses of Datasets**

To study the effect of omitting data from ancestral generations, additive genetic variance was estimated from datasets that differed in number of generations with records known. We used the complete relationship matrix (i.e., all relations were known since the start of selection) to account for inbreeding and covariances between animals.

In another set of analyses, we assumed data known for a limited number of generations after animals had been selected for several generations. Using this dataset, we compared models that differed in amount of covariances known between animals by varying the generation that was assumed to consist of unrelated, unselected and non-inbred animals.

Inbreeding and covariances among animals in a given population depend on the size of the population. We therefore have simulated two population sizes, with 40 and 400 animals per generation, respectively. Parameters for \(s\) and \(m\) were 5 and 20 for the small population and
were 20 and 200 for the large population, respectively.

The true genetic variance in generation $t$ was defined as:

$$
\sigma_a^2 = \frac{1}{n-1} \left( a_i' a_i - n \bar{a}_i^2 \right) \tag{4}
$$

where $a_i$ is a vector with true breeding values for $n$ animals in generation $t$. The variance $\sigma_a^2$ is expected to be smaller than $\sigma^2_a$ because $\sigma^2_a$ does not adjust for inbreeding and gametic disequilibrium (Bulmer, 1971). Furthermore, $\sigma^2_a$ does not adjust for covariances between animals, which is particularly present in small populations.

In a balanced hierarchical design for $n$ animals having $s$ sires and $d$ dams per sire (sires and dams unrelated), each dam having $p$ progeny, the expectation of $\sigma_a^2$ is equal to

$$
\frac{1}{4} \frac{(a - d^n)}{n-1} \text{var}(a_d) + \frac{1}{4} \frac{(a - p)}{n-1} \text{var}(a_d) + \text{var}(\phi) \tag{5}
$$

Family structures in later generations were more complicated due to covariances between sires and dams. We can write $\sigma_a^2$ more generally as $\frac{1}{n-1} a' (I - J^2) a = \frac{1}{n-1} a' Q a$, with $J$ being an $n$ by $n$ matrix with all elements equal to 1, and $I$ is the identity matrix of order $n$. With no selection, the expectation of $\sigma_a^2$ is

$$
E(\sigma_a^2) = \frac{1}{n-1} \text{tr}(Q A_{a}) \sigma_a^2 \tag{6}
$$

with $A_{a}$ being the matrix with additive genetic relationships between animals in generation $t$.

Sorensen and Kennedy (1984b) used $\sigma_a^2 / 2$ to indicate the reduction of genetic variance due to gametic disequilibrium. $F_1$ represented the average inbreeding coefficient in generation $t$. This ratio, however, does not adjust for an additional reduction due to covariances among animals. Therefore, we used $\frac{1}{2(n-1)} \text{tr}(Q A_{a}) \sigma_a^2$ as a measure of disequilibrium due to selection.

**Estimation of Genetic Variance with Restricted Maximum Likelihood**

The following model was used for analysis of the data:

$$
y = Xb + Za + e \tag{7}
$$

with

$$
E(y) = Xb
$$

$$
\text{var}(a) = \Lambda \sigma_a^2
$$

$$
\text{var}(y) = V = V_1 \sigma_a^2 + V_0 \sigma_e^2
$$

$$
= ZAZ' \sigma_a^2 + I_n \sigma_e^2
$$

Mixed model equations after absorbing fixed effects are:

$$
[Z'MZ + \alpha A^{-1}] \hat{a} = Z'My \tag{8}
$$

where $M = I - X(X'X)^{-1} X'$ and $\alpha = \sigma_e^2 / \sigma_a^2$.

We want to maximize the likelihood of the parameter vector $\tau = [\sigma_a^2 \sigma_e^2]$ in the space of error contrasts, hence maximize $(\tau | K'y)$, where $K'K = M$. The log likelihood function of $K'y$ can be written as follows (Searle, 1979; Smith and Graser, 1986):

$$
\ell(K'y) = \text{constant} + [N - \text{rank}(C)] \log \sigma_e^2
$$

$$
+ q \log \sigma_e^2 + \log |C|
$$

$$
+ y'Py / \sigma_e^2, \tag{9}
$$

where $C$ is the full rank coefficient matrix of the mixed model equations before absorption of the fixed effects, $q$ is equal to the number of random animal effects, and $P = V^{-1} - V^{-1} X (X' V^{-1} X)^{-1} X' V^{-1}$. Procedures to determine the maximum of $\ell$ have been presented by Graser et al. (1987) and Meyer (1989).

**Results**

Reduction of Genetic Variance

Due to Selection

To demonstrate the changes in genetic variance after selection in a small population, we show the mean variance at each generation averaged over 1,000 replicates. Table 1 shows the results from random mating and Table 2
TABLE 1. MEANS ($\overline{\sigma}_a$) AND VARIANCES ($\sigma_a^2$) OF TRUE ADDITIVE GENETIC VALUES, EXPECTED VARIANCE, AVERAGE INBREEDING COEFFICIENT ($F_t$) AND GAMETIC DISEQUILIBRiUM (DE) AVERAGED OVER 1,000 REPLICATES FOR FIVE GENERATIONS OF RANDOM MATING

<table>
<thead>
<tr>
<th>Generation</th>
<th>$\overline{\sigma}_a$</th>
<th>$\sigma_a^2$</th>
<th>$\frac{1}{n-1}\sigma(QA)\sigma_a^2$</th>
<th>$F_t$</th>
<th>DE$^c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-.01</td>
<td>10.03 (2.92)$^d$</td>
<td>10.00</td>
<td>.000</td>
<td>1.00</td>
</tr>
<tr>
<td>1</td>
<td>-.02</td>
<td>9.66 (2.65)</td>
<td>9.49</td>
<td>.000</td>
<td>1.02</td>
</tr>
<tr>
<td>2</td>
<td>-.04</td>
<td>9.20 (2.69)</td>
<td>9.17</td>
<td>.016</td>
<td>1.01</td>
</tr>
<tr>
<td>3</td>
<td>-.01</td>
<td>8.91 (2.53)</td>
<td>8.93</td>
<td>.037</td>
<td>1.00</td>
</tr>
<tr>
<td>4</td>
<td>-.03</td>
<td>8.71 (2.57)</td>
<td>8.69</td>
<td>.059</td>
<td>1.00</td>
</tr>
<tr>
<td>5</td>
<td>-.01</td>
<td>8.46 (2.44)</td>
<td>8.47</td>
<td>.083</td>
<td>1.00</td>
</tr>
</tbody>
</table>

$^a$Initial genetic variance was 10 and each generation contained 40 animals.
$^b$Expected variance after adjusting for covariances and inbreeding.
$^c$Disequilibrium computed as $\sigma_n^2 \cdot \frac{1}{n-1}\sigma(QA)\sigma_n^2$.
$^d$Empirical standard deviation.

shows these from selection in a population of size 40 (=2m).

Mean additive genetic values were close to zero for all generations with random mating. The additive genetic variance declined due to the establishment of covariances between animals and an increase of the average inbreeding coefficient ($F_t$). The expected genetic variance according to [6] has been adjusted for both covariances and inbreeding. Deviations of the coefficient for disequilibrium (DE) from unity represented gametic disequilibrium. In case of no selection, this deviation is small and generated by chance only. Gametic equilibrium (or selection disequilibrium) also has been referred to as linkage disequilibrium, which we consider a confusing term because loci are assumed to be unlinked.

Table 2 shows the change of means and variance of additive genetic values for 10 generations with selection in the small population. There a clear response to selection and the reduction of $\sigma_a^2$ was significant. As expected, the average inbreeding was somewhat higher with selection. The genetic variance decreased more compared to the situation with random mating. An additional decline of genetic variance was due to the establishment of disequilibrium. After some generations of selection, the coefficient for disequilibrium (DE) stabilized at a constant value because new disequilibrium and recombination offset each other.

For infinite population size, the reduction of variance after truncation selection is equal to $k = i/(1-x)$, where $x$ is the truncation point and $i$ is the intensity of selection. Becker (1975) gives $i = 1.27$ and $k = .759$ for a selected portion of 25%. Selection intensities for small sample sizes can be derived from order statistics (e.g., $i = 1.2145$ when selecting 5 out of 20 (Beyer, 1968; Becker, 1975). The effective selection intensity is reduced further when selection is on correlated variables. The first generation of our population consisted of five half-sib families and 20 full-sib families with intraclass correlation equal to $\frac{1}{4}h^2$ and $\frac{1}{2}h^2$, respectively. From the selection response we can derive the obtained selection intensity to be equal to 1.12. This is in accordance with Table 1 from Hill (1977).

We used order statistics to compute the variance among a selected group as well. The variance among the five highest ranking out of 20 was 20.7% of the variance before selection, giving $k = .795$. Apparently, the reduction of variance after selection in small samples is higher compared to infinite population size. However, selection in small populations also usually involves selection on correlated variables because animals are more related to each other; this gives a smaller reduction in genetic variance. Because values for DE had been corrected for covariances and inbreeding, we used $\sigma^2 = DE\sigma^2_0$ in Bulmer's (1971) formula for an infinite population,

$$\sigma_n^2 = \frac{1}{4}(1-k^2)\sigma_{h_n}^2 + \frac{1}{4}\sigma_{h_{n-1}}^2 + \frac{1}{2}\sigma_{0n}^2$$

to derive empirically a k-value of about .6 for $t = 2$ and of about .5 in later generations.
TABLE 2. MEANS ($\bar{s}$) AND VARIANCES ($\sigma^2$) OF TRUE ADDITIVE GENETIC VALUES, EXPECTED VARIANCE, AVERAGE INBREEDING COEFFICIENT ($\bar{F}$) AND SELECTION DISEQUILIBRIUM (DE) AVERAGED OVER 1,000 REPLICATES FOR 10 GENERATIONS OF SELECTION IN SMALL POPULATIONS\textsuperscript{ab}

<table>
<thead>
<tr>
<th>Generation</th>
<th>$\bar{s}$</th>
<th>$\sigma^2$</th>
<th>$\frac{1}{n-1}\text{tr}(QA)\sigma^2$</th>
<th>$\bar{F}$</th>
<th>DE\textsuperscript{d}</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>.05</td>
<td>9.96 (2.83)\textsuperscript{b}</td>
<td>10.00</td>
<td>.000</td>
<td>1.00</td>
</tr>
<tr>
<td>1</td>
<td>.06</td>
<td>9.37 (2.71)</td>
<td>9.49</td>
<td>.000</td>
<td>.99</td>
</tr>
<tr>
<td>2</td>
<td>1.31</td>
<td>8.47 (2.34)</td>
<td>9.14</td>
<td>.017</td>
<td>.93</td>
</tr>
<tr>
<td>3</td>
<td>2.45</td>
<td>8.05 (2.27)</td>
<td>8.85</td>
<td>.040</td>
<td>.91</td>
</tr>
<tr>
<td>4</td>
<td>3.55</td>
<td>7.83 (2.07)</td>
<td>8.64</td>
<td>.065</td>
<td>.91</td>
</tr>
<tr>
<td>5</td>
<td>4.65</td>
<td>7.59 (2.03)</td>
<td>8.39</td>
<td>.089</td>
<td>.90</td>
</tr>
<tr>
<td>6</td>
<td>5.76</td>
<td>7.32 (1.79)</td>
<td>8.17</td>
<td>.113</td>
<td>.90</td>
</tr>
<tr>
<td>7</td>
<td>6.81</td>
<td>7.10 (1.79)</td>
<td>7.98</td>
<td>.137</td>
<td>.89</td>
</tr>
<tr>
<td>8</td>
<td>7.86</td>
<td>6.97 (1.84)</td>
<td>7.76</td>
<td>.159</td>
<td>.90</td>
</tr>
<tr>
<td>9</td>
<td>8.86</td>
<td>6.80 (1.83)</td>
<td>7.55</td>
<td>.181</td>
<td>.90</td>
</tr>
<tr>
<td>10</td>
<td>9.84</td>
<td>6.72 (1.86)</td>
<td>7.37</td>
<td>.203</td>
<td>.91</td>
</tr>
</tbody>
</table>

\textsuperscript{a}Initial genetic variance was 10 and each generation contained 40 animals.
\textsuperscript{b}Five out of twenty males were selected each generation.
\textsuperscript{c}Expected variance after adjusting for covariances and inbreeding.
\textsuperscript{d}Disequilibrium computed as $\sigma^2 = \frac{1}{n-1}\text{tr}(QA)\sigma^2$.
\textsuperscript{e}Empirical standard deviation from 1,000 replicates.

\textit{Estimation of Genetic Variance When Data from Selected Animals are Missing}

Genetic variance was estimated using different models and datasets. Average true genetic variances are given for 1,000 replications for the small population ($2m = 40$) in Table 2 and for 20 replications for the large population ($2m = 400$) in Table 3. Selection response and effect of gametic disequilibrium on genetic variance was higher in the large population, due to the higher selection intensity. However, a reduction of genetic variance was smaller because of less inbreeding in this population (Table 3).

Estimated additive genetic variances by subsequently omitting data from parental generations are given for the two populations in Table 4. Some of the replicates for the small population did not converge to a solution and the estimate for $\sigma^2$ approached 0. Those

TABLE 3. MEANS ($\bar{s}$) AND VARIANCES ($\sigma^2$) OF TRUE ADDITIVE GENETIC VALUES, EXPECTED VARIANCE, AVERAGE INBREEDING COEFFICIENT ($\bar{F}$) AND SELECTION DISEQUILIBRIUM (DE) AVERAGED OVER 20 REPLICATES FOR 5 GENERATIONS OF SELECTION IN LARGE POPULATIONS\textsuperscript{ab}

<table>
<thead>
<tr>
<th>Generation</th>
<th>$\bar{s}$</th>
<th>$\sigma^2$</th>
<th>$\frac{1}{n-1}\text{tr}(QA)\sigma^2$</th>
<th>$\bar{F}$</th>
<th>DE\textsuperscript{d}</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.05</td>
<td>9.82 (1.19)\textsuperscript{b}</td>
<td>9.88</td>
<td>.000</td>
<td>.99</td>
</tr>
<tr>
<td>2</td>
<td>1.90</td>
<td>8.92 (.78)</td>
<td>9.79</td>
<td>.005</td>
<td>.91</td>
</tr>
<tr>
<td>3</td>
<td>3.72</td>
<td>8.55 (.76)</td>
<td>9.72</td>
<td>.013</td>
<td>.88</td>
</tr>
<tr>
<td>4</td>
<td>5.47</td>
<td>8.19 (.88)</td>
<td>9.61</td>
<td>.020</td>
<td>.85</td>
</tr>
<tr>
<td>5</td>
<td>7.14</td>
<td>8.12 (.79)</td>
<td>9.54</td>
<td>.027</td>
<td>.83</td>
</tr>
</tbody>
</table>

\textsuperscript{a}Initial genetic variance was 10 and each generation contained 400 animals.
\textsuperscript{b}Twenty out of 200 males were selected each generation.
\textsuperscript{c}Expected variance after adjusting for covariances and inbreeding.
\textsuperscript{d}Disequilibrium computed as $\sigma^2 = \frac{1}{n-1}\text{tr}(QA)\sigma^2$.
\textsuperscript{e}Empirical standard deviation from 20 replicates.
replicates were omitted for calculating the mean result.

The estimate of additive genetic variance using REML was close to the initial value when all data and the complete relationships matrix were used (Table 4). This agrees with the result of REML accounting for selection using the complete mixed model (Sorensen and Kennedy, 1984b; Gianola and Fernando, 1986).

Subsequently omitting data in the small population from generations 1, 1 to 2 and 1 to 3 gave estimates for \( \sigma^2_a \) equal to 9.49, 9.54, and 9.60, respectively (Table 4). This suggests that the relationships matrix accounted for most of the selection, even though the analysis did not include records on which selection decisions had been based. The empirical standard deviation of the estimates increased considerably, however, when data were omitted from the analysis. Omitting data from selected generations in the large population had more effect on the mean estimated genetic variance; estimates of \( \sigma^2_a \) decreased more when data from more previous generations were omitted. Decrease of mean estimate also was larger when considered in proportion to the higher coefficient of disequilibrium for the large population.

Table 5 shows the effect of assuming generation 5 as the base population (unselected, unrelated and non-inbred). In this case, the assumed base population had a smaller variance than twice the variance generated by Mendelian sampling. The genetic variance was estimated omitting data from subsequent generations. Omitting data had little effect on the mean estimate. Similar results were found when generation 6 was assumed to be the base generation. Adding data from more subsequent generations affected estimate of genetic variance in terms of standard error, but not in terms of the mean. Notice that with data from generations 6 and 7, and considering generation 6 as the base, an estimate was obtained, which was higher than \( \sigma^2_a \). This is not in agreement with the result of Sorensen and Kennedy (1984b).

### Including Additional Relationships Between Base Animals

Usually in data analysis, an arbitrary generation is treated as consisting of unselected, unrelated and non-inbred base animals. Results from Table 5 suggest that bias from prior selection could be (partly) removed when

### TABLE 4. ESTIMATED GENETIC VARIANCE (\( \sigma^2_a \)) AFTER OMITTING DATA FROM AN INCREASING NUMBER OF SELECTED GENERATIONS BUT INCLUDING THE COMPLETE RELATIONSHIP MATRIX

<table>
<thead>
<tr>
<th>Data used from</th>
<th>( \sigma^2_a ) (small pop.(^b))</th>
<th>No.</th>
<th>( \sigma^2_a ) (large pop.(^b))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gen. 1–5 (all data)</td>
<td>9.81 (2.78)(^c)</td>
<td>1,000</td>
<td>10.09 (.81)(^d)</td>
</tr>
<tr>
<td>Gen. 2–5</td>
<td>9.49 (3.22)</td>
<td>997</td>
<td>9.90 (.83)</td>
</tr>
<tr>
<td>Gen. 3–5</td>
<td>9.54 (3.83)</td>
<td>980</td>
<td>9.58 (1.17)</td>
</tr>
<tr>
<td>Gen. 4–5</td>
<td>9.60 (4.76)</td>
<td>901</td>
<td>8.93 (1.94)</td>
</tr>
</tbody>
</table>

\(^a\)Five out of twenty males were selected each generation.
\(^b\)Twenty out of 200 males were selected each generation.
\(^c\)Empirical standard deviations from n replicates.
\(^d\)Empirical standard deviations from 20 replicates.

### TABLE 5. ESTIMATED GENETIC VARIANCE (\( \sigma^2_a \)) AFTER OMITTING DATA FROM AN INCREASING NUMBER OF SELECTED GENERATIONS FOR A GIVEN SET OF COVARIANCES BETWEEN ANIMALS\(^a\)

<table>
<thead>
<tr>
<th>Data used from</th>
<th>Base generation(^b)</th>
<th>( \sigma^2_a )</th>
<th>No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gen. 9–10</td>
<td>5</td>
<td>8.56 (4.29)(^c)</td>
<td>855</td>
</tr>
<tr>
<td>Gen. 8–10</td>
<td>5</td>
<td>8.15 (3.42)</td>
<td>973</td>
</tr>
<tr>
<td>Gen. 7–10</td>
<td>5</td>
<td>8.20 (2.84)</td>
<td>990</td>
</tr>
<tr>
<td>Gen. 6–10</td>
<td>5</td>
<td>8.30 (2.42)</td>
<td>999</td>
</tr>
<tr>
<td>Gen. 6–7</td>
<td>6</td>
<td>7.81 (3.85)</td>
<td>918</td>
</tr>
<tr>
<td>Gen. 6–8</td>
<td>6</td>
<td>7.78 (3.21)</td>
<td>982</td>
</tr>
<tr>
<td>Gen. 6–9</td>
<td>6</td>
<td>7.80 (2.61)</td>
<td>994</td>
</tr>
<tr>
<td>Gen. 6–10</td>
<td>6</td>
<td>7.92 (2.31)</td>
<td>1,000</td>
</tr>
</tbody>
</table>

\(^a\)Five out of twenty males were selected each generation.
\(^b\)Generation assumed to consist of unrelated, non-inbred and unselected animals.
\(^c\)Empirical standard deviation from n replicates.
TABLE 6. ESTIMATED GENETIC VARIANCE ($\theta_a^2$) USING DATA FROM TWO SELECTED GENERATIONS AND INCLUDING THE RELATIONSHIPS GENERATED FROM A VARIOUS NUMBER OF GENERATIONS

<table>
<thead>
<tr>
<th>Data from</th>
<th>Base gen.</th>
<th>$\theta_a^2$</th>
<th>Bias, %</th>
<th>No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gen. 4-5</td>
<td>3</td>
<td>8.71 (4.33)</td>
<td>-7.3</td>
<td>924</td>
</tr>
<tr>
<td>Gen. 4-5</td>
<td>2</td>
<td>9.05 (4.56)</td>
<td>-5.9</td>
<td>892</td>
</tr>
<tr>
<td>Gen. 4-5</td>
<td>1</td>
<td>9.34 (4.68)</td>
<td>-5.6</td>
<td>896</td>
</tr>
<tr>
<td>Gen. 4-5</td>
<td>0</td>
<td>9.60 (4.76)</td>
<td>-4.6</td>
<td>901</td>
</tr>
<tr>
<td>Gen. 9-10</td>
<td>8</td>
<td>7.76 (3.97)</td>
<td>-7.4</td>
<td>903</td>
</tr>
<tr>
<td>Gen. 9-10</td>
<td>7</td>
<td>8.07 (4.08)</td>
<td>-7.1</td>
<td>864</td>
</tr>
<tr>
<td>Gen. 9-10</td>
<td>6</td>
<td>8.33 (4.19)</td>
<td>-7.5</td>
<td>859</td>
</tr>
<tr>
<td>Gen. 9-10</td>
<td>5</td>
<td>8.56 (4.29)</td>
<td>-6.3</td>
<td>855</td>
</tr>
</tbody>
</table>

*Five out of twenty males were selected each generation.

*Generation assumed to consist of unrelated, non-inbred and unselected animals.

*Reference values computed as $\theta_a^2$ for the case of no selection.

*Empirical standard deviation from n replicates.

relationships between these assumed base animals are included in the model. The effect of including relationships established in earlier generations is shown in Tables 6 and 7 for the small and the large population, respectively. Data were used from animals of two generations and relationships known from an increasing number of previous generations were included.

Estimates of genetic variance were higher than the true variance of the assumed base population. Estimates were compared with values that were computed using the same model, but with data obtained for the case of no selection of males. Results indicate that adding additional relationships between base animals accounted for covariances and inbreeding. Table 6 shows that estimates generally are only slightly less biased by disequilibrium from selection, when additional relationships between the assumed base animals are included.

In the large population, there was less accumulation of inbreeding, and the effect of including more relationships was smaller (Table 7). However, the bias from gametic disequilibrium was larger in the small population. This bias was reduced only slightly by including additional relationships between base animals.

### Discussion

We considered selection for metric trait, which is equally affected by many unlinked loci. The change of genetic variance due to changes in gene frequencies is small and can be ignored in such an infinitesimal model.

Changes due to inbreeding are irreversible. However, the coefficient of inbreeding is only a relative measure (i.e., a base population is a population that by definition has an average inbreeding coefficient of zero) (Falconer, 1989). In practice, choosing a generation is arbitrary, and accounting for inbreeding in previous generations does not provide estimates that are better able to predict future genetic gain because newly generated Mendelian sampling variance will be reduced consistently through inbreeding.

Bulmer (1971) has pointed out that gametic disequilibrium vanishes after selection is ceased. Moreover, the variance that is generated at each generation by recombination is not affected by disequilibrium. Estimates of genes...
Genetic variance therefore should be corrected for the effect of disequilibrium.

Robertson (1977) indicated that estimates of genetic variance based on half- and full-sib analysis will be biased due to a reduction of genetic variance among selected parents. Parent-offspring regression is not affected by this type of selection (Hill and Thompson, 1977). A maximum likelihood estimator, like REMI, combines information from contrasts within a generation, with parent offspring covariances and with information from contrasts between families (i.e., between base parents) (Thompson, 1977). Little is known about the weighing of information from these different sources. It was shown empirically in this study that the mean estimate of additive genetic variance was determined primarily by the generation that was assumed to be the base generation (Tables 5 to 7). This is supported by results from Table 5, which show no change in mean estimate when information on more subsequent generations is available to estimate variance from Mendelian sampling. However, this result needs theoretical verification.

REML using all relationships and all data has been shown to account for all selection. The argument that accounting for selection is possible only when all data are used (Gianola and Fernando, 1986) has been demonstrated for sequential selection within a generation, or with selection on a correlated trait (Meyer and Thompson, 1984). However, some of the genetic variance lost through selection is regained at each generation through gene segregation during meiosis (Mendelian sampling). Contributions from selected parents and those due to the Mendelian recombination can be treated as random and independently active terms. Thompson (1977) has shown the contributions of parental generations acting on the total additive genetic variance.

Let us consider to have data on generation 2 and at least one later generation. If we assume the covariance between sires and dams can be ignored, the variance in generation 2 can be written as follows:

\[
\text{var}(a_2) = \frac{1}{4} \text{var}(a_{s_1}) + \frac{1}{4} \text{var}(a_{d_1}) + \text{var}(\phi_{s_1})
\]

[10]

The variance in this generation is reduced due to selection of their sires only. Other terms are not affected by selection. Selection is on records from sires in generation 1, which is \(y_{s_1}
\]

\[
\begin{bmatrix}
\frac{1}{2} a_{s_0} \\
\frac{1}{2} a_{d_0} \\
\phi_{s_1}
\end{bmatrix}
= \begin{bmatrix}
\frac{1}{4} & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1/2
\end{bmatrix}
\sigma_a^2,
\]

\[
B = \text{cov}
\begin{bmatrix}
\frac{1}{2} a_{s_0} \\
\frac{1}{2} a_{d_0} \\
\phi_{s_1}
\end{bmatrix}
= \begin{bmatrix}
\frac{1}{4} \\
0 \\
1/2
\end{bmatrix}
\sigma_a^2
\]

\[
P = \text{var}(y_{s_1}) + (1 - k)P.
\]

From [11] it can be seen that the Bulmer effect (i.e., the reduction of genetic variance after selection) is due to the reduction of variance of each of the components of the total genetic variance, and, moreover, is due to negative covariances between these components. In genetic terms, this has been referred to as "negative covariances between loci" (Bulmer, 1971).

Even though the variance among the sires of generation 2 is reduced by \(kh^2\sigma_a^2\), the observed bias in estimation of genetic variance, using data from generation 2, was smaller than \(\frac{1}{4}kh^2\sigma_a^2\). Estimated values were 9.64 and 9.93 for the small and for the large population, respectively. A mixed model, explaining the breeding values in terms of genetic values of parental generations, reduces...
bias from the Bulmer effect, because it does not include covariances between these parental contributions. Note that the sum of the diagonal elements of \([11]\) is equal to \((1 - 0.375kh^2)s^2\), rather than \(1 - kh^2s^2\).

In the large population the bias from selection was considerably greater when data from more generations were omitted, even when the complete relationship matrix was used. After repeated cycles of selection, expression \([11]\) consists of more ancestral contributions, which are redundantly affected by selection (e.g., the sum of all diagonal elements after selection on \(y_2\) would be \((1 - 0.45kh^2)\)). Bias from disequilibrium was small in the small population, and the difference with bias in the large population was not in proportion to the difference in coefficient for disequilibrium. Apparently, the effect of selection on different genetic components is dependent on the population size (e.g., in the small population, there is a greater chance of selecting related animals) (Robertson, 1961), and the variance among Mendelian sampling terms is relatively less reduced. This tendency toward "family selection" is greater when selection was on estimated breeding values using an animal model.

Implications

Including additional relationships between the assumed base animals accounted for covariances and inbreeding, and partly for the bias from gametic disequilibrium. Relationships accounted for more bias from gametic disequilibrium when selection had been practiced for just a few generations, or when the population size was small. The mean estimate of additive genetic variance was determined primarily by which generation was assumed to be unrelated, non-inbred and unselected; having more subsequent generations with data did not affect these mean estimates. Variance components used for animal evaluation should be estimated with a model that coincides with the same base population, possibly even with different values for the component of variance between selected base animals, and for variance within families. Estimates of genetic variance to determine expected genetic progress are biased less by covariances and gametic disequilibrium when more relationships are included in the model.

Literature Cited


Thompson, R. 1977. The estimation of heritability with unbalanced data. II. Data available on more than two generations. Biometrics 33:497.