Estimation of heritability and breeding values for early egg production in laying hens from pooled data

F. Biscarini,1 H. Bovenhuis, E. D. Ellen, S. Addo, and J. A. M. van Arendonk

Animal Breeding and Genomics Centre, Wageningen University, PO Box 338, 6700 AH Wageningen, the Netherlands

ABSTRACT Under commercial conditions, data on egg production in laying hens are usually collected per cage rather than individually. In current breeding programs, genetic evaluations are, however, based on individually recorded egg production. Because commercial flocks are not maintained in single cages, this environmental difference between the breeding and commercial setting may result in a genotype × environment interaction. This study was aimed at estimating genetic parameters and predicting estimated breeding values for early egg production of laying hens by using pooled data (i.e., data from multiple bird cages) from pedigree birds housed in 4-bird cages. Using cage records, we compared 2 different methods of handling pooled data: cage sums and the assignment of cage means to individual birds, referred to as the approximate method. The 2 methods were compared by using cross-validation. Data from 3 purebred White Leghorn layer lines were used. Estimated heritability for early egg production was 0.36 when cage sums were used and 0.30 with the approximate method. The correlation of estimated breeding values between the cage sums method and the approximate method was 0.88. Cross-validation showed that the use of cage sums led to better predictions of missing phenotypes compared with the approximate method. The results of the research demonstrate that pooled data can be used in the genetic evaluation of laying hens and show that using directly pooled records (e.g., cage sums) gives better results than assigning group means to the birds of the group, thus simulating individual records.

Key words: heritability, group, laying hen, pooled data, early egg production

INTRODUCTION

In commercial egg production facilities, layers are kept in groups. In the current battery systems, hens are housed in cages of usually 4 hens each. However, in the near future, communitarian legislative developments (Council Directive 1999/74/EC) will lead to housing layers in larger groups (furnished cages, non-cage systems). In these situations, data on egg production are commonly collected for the whole group rather than individually. However because standard methods for genetic evaluation are based on individual records, selection of layers is carried out in nucleus flocks in which purebred layer lines are housed individually and selected on the basis of individually recorded traits. This constitutes a considerable cost for the breeding industry and might lead to genotype × environment interaction between the testing and commercial environments (Merks, 1989; Besbes and Ducroq, 2003). Using directly the information collected on groups of hens (pooled information) would therefore represent a financial benefit for the breeding companies and, depending on the magnitude of the genotype × environment interaction, might lead to higher genetic response at the commercial level. Theoretical work on the extension of genetic evaluation methodology by incorporating pooled information has been done with simulations by Olson et al. (2006) and with actual data by Biscarini et al. (2008). Results from their research showed that genetic evaluation based on pooled data instead of individual observations is theoretically and practically feasible. Restrictions do, however, still exist: pooled data from groups of different sizes, for instance, cannot yet be used, mainly due to software limitations (Biscarini et al., 2008). Estimates of breeding values based on pooled data are, however, less accurate than those based on individual observations (Olson et al., 2006; Biscarini et al., 2008). The application of pooled data in the genetic evaluation of traits of agricultural interest has been very limited so far.

Two main strategies to deal with pooled data in animal models have been proposed by Olson et al. (2006)
and Biscarini et al. (2008). One is to directly use the pooled record, either the sum or average of the performance of the individuals in a group, thus taking the group structure into account in the model of analysis. The other consists in assigning the average performance of a group to each member of that group and then treating the observations as if they were individual observations. This second method, known as the approximate method, is less adherent to reality because it does not account for the group structure and systematically overestimates the accuracy of the estimates (Olson et al., 2006) but is operationally easier to implement given its similarity with the standard procedure of genetic evaluation. Besides, because current software packages for genetic evaluation have been designed for individual records, it is easier to use the approximate method, especially if groups are of different sizes.

The objective of the present study was to estimate heritability and predict breeding values for early egg production based on 4-bird cage records and to compare the results of the 2 different methods (cage sums and approximate method) of analyzing pooled data. Data from 3 purebred White Leghorn layer lines were available for this study.

**MATERIALS AND METHODS**

**Data**

Data were provided by the Institut de Sélection Animale B.V. (Boxmeer, the Netherlands). For this study, the same birds were used as described in Ellen et al. (2008). The bird population consisted of 15,212 laying hens from 3 purebred lines of White Leghorn origin (lines W1, WB, and WF). Chicks were hatched in 2 batches. All 3 lines were represented in each batch. Each batch comprised chicks of 4 age classes, differing by 2 wk each. Newly hatched chicks were sexed and identified by means of wing bands applied to the right wing. They were vaccinated against Marek’s disease and infectious bronchitis. The beaks of the chicks were kept intact. During the rearing period, chicks were kept in groups of the same line and age. At 17 wk of age, hens were transported to 2 different laying facilities. Both facilities had 8 rows of cages separated by corridors, each consisting of 3 tiers (top, middle, and bottom). Hens were kept in a total of 3,803 cages with 4 hens of the same line and age each. Hens were assigned to cages at random. Due to chance, some hens in a cage were related to various degrees. The number of hens and cages per line is presented in Table 1. Hens were provided with water and a standard commercial layer diet ad libitum. The photoperiod in the laying houses was increased starting from 9 h of light per day. Each week, the photoperiod was increased by 1 h until 16 h of light per day was reached, when the hens were about 26 wk old.

The data consisted of the total eggs produced for each 4-bird cage from 17 to 24 wk of age (pooled early egg production). The number of eggs produced each day was recorded daily for each cage and was then summed over the 7-wk period (cage sums). No individual observations were available. The mean egg production of a cage was used in the approximate method, assigning it to the individual hens housed in the cage. Only cages that consisted of 4 hens during the whole period from 17 and 24 wk of age were used in the analysis (3,803 cages), and cages in which mortality occurred were excluded (412 cages).

The experimental population originated from a total of 505 sires and 2,331 dams. The same sires were used in the production of the hens that were housed in both laying facilities, but different sets of dams were used to breed the hens in the 2 facilities. On average, each sire was mated to 5 dams, resulting in 1 to 28 offspring per mating (on average 12 offspring per dam). Mating was random. Four generations of birds were extracted from the pedigree for the calculation of the additive relationship matrix.

**Data Analysis**

A preliminary ANOVA was performed to determine the systematic effects having a significant influence on early egg production. All hens from the 3 White Leghorn lines were analyzed together. The effects of laying facility, hatch week, genetic line, and position of the cage in the laying facility (combination of row and tier)
were included in the model. Hatch week was nested within laying facility. The following linear model was used for the estimation of variance components and breeding values:

\[ y_{ijklm} = \mu + L_i + P_j + S_k + H_l(k) + a_m + e_{ijklm}, \]

where \( y_{ijklm} \) is the early egg production (either cage sum or cage mean, for the pooled and approximate data analysis, respectively) of bird \( m \), of line \( i \), in cage at position \( j \), hatched at week \( l \) in laying facility \( k \); \( \mu \) is the common mean; \( L_i \) is the \( i \)th genetic line (W1, WB, or WF); \( P_j \) is the \( j \)th position of a cage in the facility (corridor \( \times \) tier) in classes; \( S_k \) is the \( k \)th laying facility; \( H_l(k) \) is the \( l \)th hatch week nested within the \( k \)th laying facility, \( a_m \) is the random genetic effect of the \( m \)th bird; and \( e_{ijklm} \) is the residual.

When using cage sums, the mixed model equations (MME) need to be modified according to Biscarini et al. (2008). The vector of observations \( y \) can be looked at as a vector of sums of individual egg productions, the incidence matrices \( X \) and \( Z \) reflect the group composition, and \( e \) is a vector of sums of the residuals. The vectors of solutions for the fixed and genetic effects, \( b \) and \( a \), are unmodified. For a couple of records, and considering only the line effect for the sake of simplicity, the MME in the case of cage sums are illustrated below. In this case, we see 2 cages with 4 hens each, the first of line WB and the second of line WF:

\[
\begin{bmatrix}
(y_1 + y_2 + y_3 + y_4) = 90 \\
(y_5 + y_6 + y_7 + y_8) = 62
\end{bmatrix}
= 
\begin{bmatrix}
4 & 0 & 4 & 0 \\
4 & 0 & 0 & 4
\end{bmatrix}
\begin{bmatrix}
\mu \\
W1 \\
WB \\
WF
\end{bmatrix}
+ 
\begin{bmatrix}
a_{s1} \\
a_{m} \\
a_{1} \\
a_{2} \\
a_{3} \\
a_{4} \\
a_{5} \\
a_{6} \\
a_{7} \\
a_{8}
\end{bmatrix}
+ 
\begin{bmatrix}
e_1 + e_2 + e_3 + e_4 \\
e_5 + e_6 + e_7 + e_8
\end{bmatrix}.
\]

The genetic variance is not affected by the modifications and is equal to \( A\sigma^2_A \), with \( A \) being the additive relationship matrix. On the contrary, because the vector \( e \) is of different nature, the residual variance is equal to \( D\sigma^2_e \), where \( D \) is a diagonal matrix with group size on the diagonal and \( \sigma^2_e \) is the individual residual variance. Because groups are of equal size (all cages have 4 hens), \( D = n \times I \) and \( R = I\sigma^2_e \), where \( I \) is the identity matrix. As a result, the ratio between the residual and additive genetic variance (\( \alpha \)) is equal to \( n \times a \) (4 \( \times \) a in this case), and the term \( R^{-1} \) can be cancelled from both sides of the MME. This implies that the residual variance estimated from cage sums is \( \sigma^2_e = n \times \sigma^2_e \) and that heritability should be calculated from variance components as \( \frac{\sigma^2_A}{\sigma^2_A + (\sigma^2_e / n)} \), where \( n \) is group size.

With the approximate method (Olson et al., 2006), the analysis is carried out exactly as in the case of the classical MME for individual observations, with the difference that the vector \( y \) of observations contains cage means attributed to each individual of the cage. Using the same example as above, the vector \( y \) would thus look as follows:

\[
\begin{bmatrix}
90/4 \\
90/4 \\
90/4 \\
62/4 \\
62/4 \\
62/4 \\
62/4
\end{bmatrix}
= 
\begin{bmatrix}
22.5 \\
22.5 \\
22.5 \\
15.5 \\
15.5 \\
15.5 \\
15.5
\end{bmatrix}.
\]

The variances of the genetic and residual effects, and their ratio \( \alpha \), are unmodified and correspond to those of the standard MME (Henderson, 1975, 1984). Residual covariances between hens in the same cage are not taken into account.

The GLM procedure of the SAS statistical software (SAS, 1996) was used for the determination of the significant fixed effects to include in the model. Variance components and breeding values for early egg production from both methods were estimated using ASReml (Gilmour et al., 2002) with a maximum likelihood approach. In the case of cage sums, the model function and \( \hat{\ } \) was used to fit multiple genetic effects per observation. Data editing and all other statistical operations were performed using the open source statistical package R.

**Cross-Validation**

The 2 methods of analyzing pooled data, the use of cage sums and the approximate method, were compared by means of cross-validation (Stone, 1974). In the cross-validation, known phenotypes are set to missing, then their values are predicted with the model and compared with the observed phenotypes. The correlation between the predicted and observed phenotypes is
From the total data set, 20% of the cages (records) at a time were randomly set to missing until each cage was once removed from the data set. Fixed effects classes were taken into account to create balanced subsets. This resulted in 5 subsets each containing 80% of the data and for which the missing phenotypes were predicted using either the exact or the approximate method. Heritability was reestimated for each subset, and phenotypes were predicted summing either the estimates for both the fixed effects and the breeding values $\hat{y}_1 = X\hat{b} + Z\hat{a}$ or only the estimates for the fixed effects $\hat{y}_2 = X\hat{b}$. Pearson correlations were calculated between the observed phenotypes and those predicted with the 2 statistical methods.

**RESULTS**

**Description of Phenotypes**

Basic statistics of the data are summarized in Table 1. The number of available cage records varied from 832 for line WF to 1,504 for line WB. Line WB had the highest early egg production, with an average of 81.5 eggs produced per cage between 17 and 24 wk: this is about twice as much as the production of line WF, which produced an average of 48.8 eggs per cage. The overall CV was 52%, with a minimum of 34% in line WB and a maximum of 61% in line WF. Being cage averages, values for the approximate method were exactly the same. Early egg production in the analyzed data set was normally distributed.

**Variance Components**

Table 2 reports variance components and heritabilities for early egg production estimated from cage sums and with the approximate method. The SE of the estimates is also reported. Estimated heritability of early egg production of laying hens is moderate, being 0.36 when using cage sums and 0.30 with the approximate method. Variance components estimated with the 2 methods differed: the ratio between additive genetic variance estimates based on cage sums and on the approximate method was 6.7, and that for estimated residual variances was 21.4. The approximated SE of the estimated heritability was 0.04 with cage sums and 0.02 with the approximate method.

**Breeding Values**

Table 3 shows the correlation between the breeding values estimated with the 2 methods. The Pearson correlation between estimated breeding values (EBV) from cage sums and from the approximate method was 0.88 for all hens with observation, 0.89 for sires with at least 10 offspring, and 0.91 for dams with at least 8 offspring. Spearman correlations were slightly lower in all 3 cases (Table 3). The ranking of birds for their genetic merit was consequently affected: when using cage sums or the approximate method, 9 of the top 20 birds were the same for all hens with observations, 14 out of 20 for sires with ≥10 offspring, and 15 out of 20 for dams with ≥8 offspring.

The accuracy of EBV was also calculated for both methods. The difference in accuracy was more marked

### Table 2. Additive genetic ($\sigma_a^2$) and residual ($\sigma_e^2$) variances, heritability ($h^2$) estimates, and SE for early egg production based on pooled observations

<table>
<thead>
<tr>
<th>Method</th>
<th>$\sigma_a^2$</th>
<th>$\sigma_e^2$</th>
<th>$h^2$</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cage sums</td>
<td>36.4</td>
<td>263.7</td>
<td>0.36</td>
<td>0.04</td>
</tr>
<tr>
<td>Approximate method</td>
<td>5.4</td>
<td>12.3</td>
<td>0.30</td>
<td>0.02</td>
</tr>
</tbody>
</table>

Heritability estimates for cage sums were calculated as $h^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2}$, where $4$ is the number of hens per cage.

### Table 3. Pearson and rank correlations between estimated breeding values (EBV) for early egg production calculated from cage sums and with the approximate method for all hens with observations, sires with more than 10 offspring, and dams with more than 8 offspring

<table>
<thead>
<tr>
<th>Correlation</th>
<th>All hens with observations</th>
<th>Sires ≥10 offspring</th>
<th>Dams ≥8 offspring</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson</td>
<td>0.88</td>
<td>0.89</td>
<td>0.91</td>
</tr>
<tr>
<td>Spearman</td>
<td>0.86</td>
<td>0.88</td>
<td>0.90</td>
</tr>
</tbody>
</table>

Correlations were based on 15,212, 166, and 996 records in the cases of all hens with observations, sires with at least 10 offspring, and dams with at least 8 offspring, respectively.
for hens with observations (0.60 vs. 0.70) and for dams (0.61 vs. 0.75) than for sires with at least 10 offspring (0.78 vs. 0.84).

**Cross-Validation**

The Pearson correlations between the observed phenotypes and their model predictions, for both the cage sums and the approximate method, are given in Table 4. When using only fixed effects to predict phenotypes, the correlations were practically the same with cage sums or the approximate method (average correlation over all subsets of 0.86 for both methods). When phenotypes were predicted adding EBV to the estimates of fixed effects, correlations between observations and predictions were consistently higher in all subsets when using cage sums in comparison to the approximate method. For the cage sums model, they ranged from 0.94 to 0.95 with an average of 0.94, whereas for the approximate method, they had an average of 0.91, ranging from 0.91 to 0.92.

**DISCUSSION**

In this study, we estimated variance components and predicted breeding values using 4-bird cage data instead of data from individual hens housed in single-bird cages. This constitutes one of the few practical applications of the theoretical work of Olson et al. (2006) and of Biscarini et al. (2008) on the use of pooled data in the genetic evaluation of farm animals. We compared 2 ways of using pooled data. One is the direct use of the pooled information, which implies that the group is the experimental unit and its total performance is the recorded observation. The other consists in apportioning the group mean to each member of the group, thus mimicking the situation in which individual observations are available. This approach is known as the approximate method (Olson et al., 2006). Cage sums are the available data and their use allows one to take into account the group structure of the data in the MME. The approximation of assigning cage means to the individual birds of each cage makes it possible to use pooled data in the usual framework for genetic evaluation of individual observations and might help work around software limitations. However, the correlation between EBV calculated with the 2 methods is well below 1, also for top sires and dams; this leads to reranking of birds and to selection of different birds for reproduction, hence the need to assess the relative quality of the 2 methods. We showed that there are reasons to consider the direct use of pooled records more appropriate.

**Cage Sums Versus the Approximate Method**

The 2 methods of analyzing pooled data, the use of cage sums and the approximate method, were used to estimate variance components and to predict breeding values, and the results were compared. Cross-validation was used to assess the validity of the proposed methods.

Heritabilities for early egg production estimated with the 2 methods were different. 0.36 with cage sums and 0.30 with the approximate method. These are not 2 independent estimates because they are both estimated from the same data but with 2 different methods. The SE of the estimate was lower for the approximate method. Accuracy of breeding values was higher with the approximate method (0.70 vs. 0.60 for hens with observations, 0.75 vs. 0.61 for dams, and 0.84 vs. 0.78 for sires). However, accuracies are obtained from the diagonal elements of the inverse of the MME (Henderson, 1975), and in the approximate method, these are not modified to account for the group structure. Therefore, with the approximate method, accuracies are the same as if every bird had an individual observation and are consequently systematically overestimated, as shown already by Olson et al. (2006). The discrepancy between true and estimated accuracy under the approximate method increases with group size (Olson et al., 2006; Biscarini et al., 2008). Therefore, we used cross-validation to compare the 2 methods in predicting phenotypes from estimated effects. When phenotypes were predicted only from fixed effects ($\hat{\gamma}_1$), there were no differences between cage sums and the approximate method. When phenotypes were predicted from the sum of the estimates of fixed effects and EBVs ($\hat{\gamma}_2$), cage sums performed consistently better than the approximate method. This indicates that the difference between the 2 methods lies in the EBV, which are better estimated by using cage sums. The results show that genetic parameters and breeding values estimated

<table>
<thead>
<tr>
<th>Method</th>
<th>$\hat{\gamma}$</th>
<th>S1</th>
<th>S2</th>
<th>S3</th>
<th>S4</th>
<th>S5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cage sums</td>
<td>$\hat{\gamma}_1$</td>
<td>0.94</td>
<td>0.95</td>
<td>0.94</td>
<td>0.94</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td>$\hat{\gamma}_2$</td>
<td>0.85</td>
<td>0.87</td>
<td>0.86</td>
<td>0.86</td>
<td>0.86</td>
</tr>
<tr>
<td>Approximate</td>
<td>$\hat{\gamma}_1$</td>
<td>0.91</td>
<td>0.92</td>
<td>0.91</td>
<td>0.92</td>
<td>0.91</td>
</tr>
<tr>
<td></td>
<td>$\hat{\gamma}_2$</td>
<td>0.86</td>
<td>0.86</td>
<td>0.86</td>
<td>0.87</td>
<td>0.85</td>
</tr>
</tbody>
</table>

1 Twenty percent of the observations were set to missing in subsets S1 to S5 and then reconstructed from the effects estimated from the model. $\hat{\gamma}_1$ and $\hat{\gamma}_2$ refer to phenotypes predicted from estimates of both systematic and genetic effects or of systematic effects only.
with the approximate method are less reliable and the
direct use of group records can therefore be regarded as
a more correct way of analyzing pooled data.

The differences observed between the 2 methods re-
side, besides the different incidence matrices \( \mathbf{X} \) and \( \mathbf{Z} \)
of the MME, in the residual variances and covariances.
When cage sums are used, the residual variance for cage
sum of 4 hens is \( \text{var}(e_1 + e_2 + e_3 + e_4) \). There might be
covariance between the residuals of birds within a cage:
this would not violate the assumption of no residual
covariance between cage sums but would affect the esti-
ated residual variance. It seems, however, reasonable
to assume that the covariance between residuals of dif-
ferent cage sums is 0. In the approximate method, the
residual variance of 4 hens in a cage is a \( 4 \times 4 \) diago-
nal matrix of the 4 residual variances [\( \text{var}(e_1), \text{var}(e_2)
\ldots \text{var}(e_4) \)]. Because all cage mates get the same phé-
notype (cage mean), the performance of a bird also
contributes to the phenotype of the cage mates. For
this reason, even if cage mates are unrelated and ob-
servations are assumed to be independent, the covari-
ance between residuals of cage mates (observations, in
this case) is not equal to 0. In the statistical analysis
presented in this work, these residual covariances be-
tween the observations from the same cage have not
been modeled. Including such residual covariances in
the model could make the results from the approximate
method more similar to those obtained with pooled re-
cords. When individual phenotypes are available, the
covariance between residuals is normally assumed to
be 0. This assumption usually holds, but this might
not be the case if the birds are cage mates. Modeling
a random cage effect could account for the covariance
among cage mates.

**Genetic Parameters**

We found a moderate level of heritability for early
egg production: estimates ranged from 0.30 with the
approximate method to 0.36 when cage sums were used.
These values do not differ substantially from heritabil-
ity estimates for early egg production based on indi-
vidual production records. Although variable, literature
estimates are in fact all of moderate to high magnitude
and fall generally in the range between 0.30 and 0.60
(Besbes et al., 1992; Wei and van der Werf, 1993, 1995;
Anang et al., 2000; Misztal and Besbes, 2000; Nurgi-
artiningisih et al., 2004). When animals are kept in
groups, as layers in cages, social interactions like com-
petition for resources, are bound to emerge (Bijma et
al., 2007a). Because we analyzed pooled data and not
individual records, we could not include the genetic as-
socative effect in the model (Bijma et al., 2007a,b).

In this study, heritability for early egg production
was estimated across lines: data of 3 different lines of
laying hens were combined in a single analysis and the
effect of line was included in the statistical model to
account for difference in genetic merit between lines.
However, heritabilities were also estimated for the indi-
vidual lines, using both cage sums and the approximate
method. With cage sums, the heritability of early egg
production was 0.23, 0.38, and 0.30 in lines W1, WB,
and WF, respectively. With the approximate method,
estimates were 0.10, 0.12, and 0.09 in lines W1, WB,
and WF, respectively. These estimates do not differ
substantially from those obtained with the combined
analysis in the case of cage sums but do differ with
the approximate method: this provides further evidence
of the unreliability of the approximate method. It is
also interesting to notice that the approximate method
seems to systematically underestimate heritability for
early egg production.

When group performances are analyzed, the residual
variance matrix \( \mathbf{R} \) is no longer \( \mathbf{L} \mathbf{e}^2 \) but \( \mathbf{D} \mathbf{e}^2 \), with \( \mathbf{D} \)
being a diagonal matrix reflecting the number of individu-
als in each group. If groups are all of equal size, then \( \mathbf{D} \)
is equal to \( n \times \mathbf{I} \), where \( n \) represents the number of
individuals in each group. This implies that the residu-
al variance estimated from group sums is expected to
be \( n \) times that estimated from individual observations.
In the present study, therefore, with cages of 4 hens
each, the residual variance should be divided by 4 to be
compared with the results of previous works on herita-
bility of early egg production. This gives a residual
variance of 37.7 for individual records, which agrees
with previous estimates, which, though variable, are in
the range of 23 to 70 (Besbes et al., 1992; Wei and van
der Werf, 1993). The expectation of the genetic vari-
ance, on the other hand, remains unmodified also in the
case of pooled data; our estimate of 36.4 is very close to
the results of Wei and van der Werf (1993), Nurgiart-
iningsih et al. (2004), and mostly in line with those of
the other works mentioned above. Biscarini et al. (2008)
estimated heritability for BW from individual and
pooled data and found good agreement between the
estimated variance components.

The variability of the estimates of variance compo-
nents and the small differences between our results and
those of other studies can be attributed to differences in
the definition of the trait and to transformation of the
data to handle lack of normality. For instance, Wei and
van der Werf (1993) defined early production as the egg
number laid between 18 and 25 wk of age, whereas in
this study, early egg production was defined as the egg
numbers from wk 17 to the end of wk 24.

Egg production traits often do not follow a normal
distribution and various mathematical techniques, such
as the Box-Cox transformation (Box and Cox, 1964),
are used to adjust for this. Our data were, however,
normally distributed and we did not apply any trans-
formation to them. Also, the use of different genetic
groups (crossbred or purebred hens) and the difference in
data structure (individual or pooled observations) can
partly account for the different estimates of variance
components. Estimates of additive genetic and residual variance with the approximate method deviated a bit more from results in the literature.

Pooled data have been used previously to estimate variance components. Simianer and Gjerde (1991) used the mean weight of samples of full-sibs to estimate variance components for BW in salmon with the minimum variance quadratic unbiased estimation methodology. Biscarini et al. (2008) adopted an approach based on modified MME using an animal model and analyzed cage sums to estimate heritability for BW in laying hens. Earlier studies on the analysis of pooled data for egg production traits in laying hens were conducted by Wei and van der Werf (1995) and by Nurgiartiningsih et al. (2004), both of whom used a sire model to estimate heritabilities from cage means of daughters. As for early egg production, Wei and van der Werf (1995) found heritabilities for eggs produced between 18 and 25 wk of age of 0.51 and 0.4 from 2 different crosses of hens; Nurgiartiningsih et al. (2004) found heritabilities for eggs produced in the first month of lay of 0.32 and 0.38 in 2 lines of White Leghorn hens.

**Accuracies of Selection**

In this study, we estimated breeding values for early egg production for 15,212 hens. However, these EBV were not based on as many individual observations but, given that only pooled data were available, they were based on 3,803 cage records. This reduced number of available records, together with the fact that pooled information, being either a cage mean or a cage sum, is a poorer source of information, leads inevitably to a lower accuracy of estimates, as has been shown already by Olson et al. (2006) and by Biscarini et al. (2008). The same authors also showed that loss of accuracy increases with group size. However, increasing the number of available records, though pooled, will lead to higher accuracy of EBV. In the present work, accuracies of 0.60, 0.78, and 0.61 were obtained for hens with records, sires, and dams, respectively. These results can be compared with those obtained by Biscarini et al. (2008) for total egg production from 371 cage records. They calculated an accuracy of 0.42, 0.50, and 0.45 for hens with records, sires, and dams, respectively. It can be seen that increasing the number of records improves the accuracy of the estimates. More precisely, with 10 times as many records (371 and 3,803, in the 2 studies), the accuracy increased 45% on average and by 55% for sires.

Besides, as mentioned by Biscarini et al. (2008), the reduction in accuracy due to the use of pooled data should also be interpreted in the context of direct versus indirect selection. In the case of laying hens, group housing at the commercial level and individual housing in the test station are 2 different environments; therefore, the genetic correlation between EBV estimated in the 2 situations is relevant for response to selection. For the same selection intensity, the ratio of the selection response for direct and indirect selection is a function of the accuracies for both situations and the genetic correlation between both traits (Falconer, 1989).

The approximate method does not take into account that the observations used in the coefficient matrix of the MME are records from a group of animals. Therefore, the accuracy of prediction is systematically overestimated with this method (Olson et al., 2006).

**Use of Pooled Data in Selection**

This paper illustrates a practical application of the use of pooled data in the genetic evaluation of laying hens. The results showed that the direct use of pooled data (e.g., cage sums) is preferable to simulating the availability of individual records by assigning the group mean to each of the birds in the group (approximate method). Modifying the MME to analyze pooled data gives in fact more accurate estimates of genetic parameters and genetic values. In a comparison of results from individual and pooled data, Biscarini et al. (2008) showed the potential for the use of pooled data in genetic evaluations of laying hens.

The use of pooled data entails a certain loss of accuracy that, however, can be compensated for by using a higher number of records. In addition, group composition plays a role: pooled observations from groups of closely related animals (e.g., full-sibs) are more difficult to attribute to the individual genotypes contributing to the observation and make the statistical analysis less accurate (Biscarini et al., 2008). This loss of accuracy, however, might not be so negative in light of the potential response to selection under the test versus commercial conditions. Moreover, the use of pooled data offers some advantages: they are often easier and cheaper to collect, they might in some cases be the only data available, and they may sometimes better reflect the commercial environment under which the offspring will be kept, thus avoiding the potential consequences of × environment interactions. Therefore, selecting individuals based on group performance may be very convenient in many situations. It will make genetic evaluation of farm animals more flexible and possibly more accurate by including information collected under commercial conditions, in which animals are kept in groups, and offers interesting possibilities for breeding companies.

**ACKNOWLEDGMENTS**

This work was conducted as part of the SABRETRAIN Project, funded by the Marie Curie Host Fellowships for Early Stage Research Training, as part of the 6th Framework Programme of the European Commission. Data were collected for the project “The genetics of robustness in laying hens,” which was a collaboration between Wageningen University and Hendrix Genetics, and financially supported by Senter Novem.
REFERENCES