

Estimation of variance components and prediction of breeding values using pooled data¹

F. Biscarini,² H. Bovenhuis, and J. A. M. van Arendonk

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

ABSTRACT: Data on performance of animals are, in several situations, collected at the group rather than individual level. Genetic evaluations in farm animals, however, are based on phenotypic information collected at the individual level. Therefore, it would be very attractive to extend genetic evaluations by incorporating information collected at the group level. In this paper we show the use of data collected at the group level for the estimation of variance components and the prediction of breeding values. We outline a general procedure that can be applied in different farm animal species. In the present work this procedure was applied to BW, for which pooled, as well as individual, observations were available, thus allowing for a comparison of the estimates, and to egg production, for which only pooled

data were available. For BW at 19 and 27 wk the estimated heritabilities based on individual observations were very similar to those based on pooled observations. For BW at 43 and 51 wk, heritability estimates based on individual and pooled data were different, which can be caused by the emergence of competition effects. The accuracy of EBV predicted from pooled observations was about 70 to 80% of the accuracy of EBV predicted from individual observations. This result quantifies the loss deriving from the use of pooled instead of individual observations. Results show that estimation of variance components and breeding values from pooled data instead of individual observations is theoretically and practically feasible.

Key words: cage performance, genetic evaluation, group, laying hen

©2008 American Society of Animal Science. All rights reserved.

J. Anim. Sci. 2008. 96:2845–2852
doi:10.2527/jas.2007-0757

INTRODUCTION

Data on performance of animals are, in several situations, collected at a group rather than individual level. The reason for this might be that animals are housed in groups and collection of information at individual level under these circumstances is too expensive or too complicated. For instance, information on feed intake and feed conversion rate is in many species measured at the group level. Also egg production of laying hens is recorded at group level under commercial conditions. This is especially relevant for species that are routinely housed in cages, tanks, or pens, such as poultry, fishes, and pigs, where data can be recorded for the whole group.

Moreover, in current breeding schemes, pure lines in nucleus breeding herds are usually housed individually and selected on the basis of individually recorded traits, whereas at the commercial level their offspring are kept in groups. Individual housing at the nucleus breeding herds is not only more expensive, but might also give rise to genotype × environment interaction (e.g., Besbes and Ducroq, 2003) that can reduce the response to selection at the commercial level. Therefore, it would be very attractive to extend genetic evaluations of farm animals by incorporating information collected at the group level. This was recognized by Olson et al. (2006) who used simulations to calculate breeding values and accuracies for pooled and individual observations.

The objective of this paper was to develop a method for estimating genetic parameters and predicting breeding values from information collected on groups of animals (pooled information) and compare its relative efficiency with the use of information collected on individual animals. For our analysis, data on individual and pooled BW of laying hens were available. The methodology was also applied to the trait egg production, for which no individual measurements were available.

¹This 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.

²Corresponding author: filippo.biscarini@wur.nl

Received November 27, 2007.

Accepted May 23, 2008.

Table 1. Distribution per origin and line of the hens and cages used in the study

Breed	Line	Hens, n	Cages, n
Rhode Island Red	B1	196	49
	B2	200	50
	B3	172	43
	BA	188	47
	BB	196	49
	BE	188	47
	Total	1,140	285
White Leghorn	W1	180	45
	WA	196	49
	WA	188	47
	WC	200	50
	WD	152	38
	WF	188	47
	Total	1,104	276

MATERIALS AND METHODS

Description of Data

The experiment was carried out by licensed and authorized personnel under approval of Institut de Sélection Animale B.V.

The animal population used in this study consisted of about 2,500 laying hens from 12 distinct genetic lines housed together in a single stable of an experimental farm for a laying period of 51 wk (wk 19 to 69 in terms of age of the hens), from June 2004 to June 2005. These lines were either of Rhode Island Red type (**RIR**; laying brown eggs) or of White Leghorn type (**WL**; laying white eggs).

Both during the rearing and the experimental period, the hens were housed in a total of 641 cages of different size (from 2 to 5 birds); only cages with 4 hens were used in this study (~560 cages). About one-half of the cages were composed of full sibs, and the other one-half had a random composition. The number of hens and cages per line and origin is presented in Table 1.

The hens had intact beaks and routine vaccinations against Marek's disease, Newcastle disease, infectious bronchitis, infectious bursal disease, fowl pox, and avian encephalomyelitis were applied during rearing. During the experiment, feed and water were available ad libitum. From the beginning of the experiment (at 19 wk of age) until 42 wk of age hens were fed a standard commercial phase 1 diet (159 g/kg of CP, 43 g/kg of crude fiber, and 11.17 MJ of ME/kg); from 42 wk onwards, until the end of the experiment, a standard commercial phase 2 diet (152 g/kg of CP, 47.0 g/kg of crude fiber, and 11.01 MJ of ME/kg) was given. Wing bands allowed individual identification of the hens. Hens arrived at the laying facility at 17 wk of age and were kept on a 9L:15D light scheme (light from 0700 until 1600 h), where **L** stands for light and **D** for dark-

ness. After 1 wk, the light period was increased by 30 min, starting at 0630 h. Thereafter, the light period was increased approximately 10 min per day. From 30 wk onwards hens received light from 0000 until 1600 h (16L:8D).

The data consisted of individual records of the BW of the hens at 19 (**BW19**, onset of the egg production), 27 (**BW27**), 43 (**BW43**), and 51 (**BW51**) wk of age. These were then pooled (i.e., summed by cage) to derive the group observations, thus allowing for comparison between genetic evaluations from individual and pooled observations. Because of software limitations, only cages with 4 hens were used. As a consequence of mortality, the number of cages with 4 hens decreased over time. In addition to BW, egg production of the entire laying period (wk 19 to 69) of 550 cages was used. For this trait there were no individual observations, but only pooled data were available.

The 4 BW traits were all normally distributed. To account for non-normality of egg production, the power transformation method of Box and Cox (1964) was applied to produce standardized variates, $z(t)$, according to the following formula:

$$z(t) = \frac{y^t - 1}{tG_y^{t-1}},$$

where y is the original observation, G_y the geometric mean of the data, and t is the parameter by which data are normalized. The value of $t = 5$, empirically derived, was used to approximate a normal distribution.

From the pedigree files, 4 generations of ancestors were extracted for the genetic analysis; sires had 1 to 30 daughters, whereas dams had 1 to 5 daughters. This is consistent with a hierarchical structure in which a rooster is mated to more females, whereas a female is only mated to 1 male.

Theoretical Background

Following the work of Olson et al. (2006), we set up the mixed model equations (**MME**) for the prediction of breeding values and the estimation of variance components using group observations.

In the case of individual records, the model, in matrix notation, is

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{e}, \quad [1]$$

with $\mathbf{y}$ being the vector of observations, $\mathbf{b}$ and $\mathbf{a}$ the vectors of fixed and random effects, with their respective incidence matrices $\mathbf{X}$ and $\mathbf{Z}$, and $\mathbf{e}$ a vector of random residuals. $\text{Var}(\mathbf{y}) = \text{Var}(\mathbf{a}) + \text{Var}(\mathbf{e})$, which, after substitution, becomes $\text{Var}(\mathbf{y}) = \mathbf{ZGZ}' + \mathbf{R}$, with $\mathbf{G} = \mathbf{A}\sigma_a^2$ and $\mathbf{R} = \mathbf{I}\sigma_e^2$.

Solutions for fixed effects and BLUP of $\mathbf{a}$ (i.e., $\hat{\mathbf{a}}$) can be then obtained by solving the usual MME (Henderson, 1975, 1984):

$$\begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{X} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \mathbf{A}^{-1}\alpha \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{bmatrix}.$$

In a numerical example, we have 4 founding animals and measured individual records on 6 of their progeny, 2 from group A, 2 from group B, and 2 from group C; α , the ratio between residual and genetic variance, is 2 (h^2 of 0.33). The pedigree and data for the example are in Table 2.

Only the fixed effect of the mean is considered, and the matrices for the MME have the usual appearance:

$$\mathbf{y} = \begin{bmatrix} 10 \\ 15 \\ 12 \\ 9 \\ 11 \\ 16 \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \mathbf{b} = [\bar{\mu}],$$

$$\mathbf{Z} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

$$\mathbf{a} = \begin{bmatrix} \hat{\mathbf{a}}_1 \\ \hat{\mathbf{a}}_2 \\ \hat{\mathbf{a}}_3 \\ \hat{\mathbf{a}}_4 \\ \hat{\mathbf{a}}_5 \\ \hat{\mathbf{a}}_6 \\ \hat{\mathbf{a}}_7 \\ \hat{\mathbf{a}}_8 \\ \hat{\mathbf{a}}_9 \\ \hat{\mathbf{a}}_{10} \end{bmatrix}, \mathbf{e} = \begin{bmatrix} \hat{\mathbf{e}}_1 \\ \hat{\mathbf{e}}_2 \\ \hat{\mathbf{e}}_3 \\ \hat{\mathbf{e}}_4 \\ \hat{\mathbf{e}}_5 \\ \hat{\mathbf{e}}_6 \\ \hat{\mathbf{e}}_7 \\ \hat{\mathbf{e}}_8 \\ \hat{\mathbf{e}}_9 \\ \hat{\mathbf{e}}_{10} \end{bmatrix}.$$

In this case, $\text{Var}(\mathbf{a}) = \mathbf{G} = \mathbf{A}\sigma_a^2$, where $\mathbf{A}$ is the additive relationship matrix, and $\text{Var}(\mathbf{e}) = \mathbf{R} = \mathbf{I}\sigma_e^2$, where $\mathbf{I}$ is an identity matrix (i.e., assuming that there are no residual correlations between animals of the same group). Taking the diagonal elements of the inverse of the MME (Henderson, 1975), the accuracies of predictions can be calculated with the following formula:

Table 2. Pedigree and data example

Animal	Sire	Dam	Trait	Group
1	0	0	—	
2	0	0	—	
3	0	0	—	
4	0	0	—	
5	1	2	10	A
6	1	2	15	A
7	1	2	12	B
8	3	2	9	B
9	3	4	11	C
10	1	4	16	C

$$r = \sqrt{1 - d_i\alpha},$$

where d_i is the i th diagonal element of the MME^{-1} and α is the ratio of variances $\left(\frac{\sigma_e^2}{\sigma_a^2}\right)$. The BLUP solutions

for the mean and the breeding values of the 10 animals, obtained by direct inversion of the MME, are shown in Table 3.

Suppose the 6 phenotypes are not recorded individually, but pooled by group. In this case the observations are the group sums and the model has to be modified accordingly (the use of group sums or group means as pooled observations proved to be equivalent when taking into account the different role of the number of observations n in the 2 cases; results not shown):

$$\mathbf{y}^* = \mathbf{X}^* \mathbf{b} + \mathbf{Z}^* \mathbf{a} + \mathbf{e}^*, \quad [2]$$

where the * symbol indicates the modified elements of the equation. The MME therefore are

$$\mathbf{y}^* = \begin{bmatrix} 10 + 15 \\ 12 + 9 \\ 11 + 16 \end{bmatrix} = \begin{bmatrix} 25 \\ 21 \\ 27 \end{bmatrix}, \mathbf{X}^* = \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix},$$

$$\mathbf{Z}^* = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix},$$

$$\mathbf{e}^* = \begin{bmatrix} \mathbf{e}_1 + \mathbf{e}_2 \\ \mathbf{e}_3 + \mathbf{e}_4 \\ \mathbf{e}_5 + \mathbf{e}_6 \end{bmatrix}.$$

Groups are now the experimental units (the observations), and, therefore, some of the elements of the MME must consequently be modified. The matrices $\mathbf{X}^*$ and $\mathbf{Z}^*$ now reflect the composition of each group: in the $\mathbf{X}$ matrix, the number of times each fixed effect is present

in each group is given, whereas in the matrix **Z** in each row there are as many 1s as animals in the group. The vector **y*** consists of the sums of the individual observations and **e*** is the sum of the residuals. Vector **a** remains unchanged, and in this example, with only a general mean, the vector **b** also remains unchanged. The variance of the genetic effect is also unmodified and equal to $\mathbf{A}\sigma_a^2$; on the other hand, because $\mathbf{e}^* \neq \mathbf{e}$, the variance of the residuals is now $\text{Var}(\mathbf{e}^*) = \mathbf{R}^* = \mathbf{D}\sigma_e^2$: assuming that there are no residual correlations between animals of the same group, **D** is not an identity matrix but a diagonal matrix with elements n_j representing the number of animals that contributed to the j th pooled observation. This affects also the value of α , which is now n times that for individual observations. In our example, where we have 2 observations in each group, $\alpha^* = 2 \times \alpha = 4$. If groups are not of equal size, the relation $\alpha^* = n \times \alpha$ does not hold and $\mathbf{R}^{*-1}$ must be used in the MME.

With this approach, BLUP of the fixed and random effects can be obtained in the very same way from the pooled observations, setting up the MME as previously outlined and solving them by direct inversion or iteratively. The solutions for the animals of the numerical example are reported in Table 3. For some animals EBV from pooled data differ considerably from EBV based on individual observations. The example illustrates that for instance full sib individuals 5 and 6, who are in the same group (A), have the same EBV when estimates are based on pooled observations. Individual 7, which is also a full sib of individuals 5 and 6, but is in another group (B), has an EBV different from 5 and 6. Following the procedure outlined by Mrode (2005), the solutions can be partitioned into parent averages, yield deviations, and progeny contributions; this helps in explaining the differences between the 2 models. The main reason behind the observed differences is that the pooling of records leads to loss of information and reduced variation: in the case of pooled data, deviations from the mean are, in general, smaller than those computed from individual records. Animal 6, for

instance, has an individual record of 15, which is a 20% deviation from the overall mean (12.25); its pooled observation is 25, which deviates only 2% from the overall mean of the pooled data (24.5). Its weighted yield deviation is, therefore, 0.575 in the individual model and 0.0616 in the pooled model, and this explains the different EBV in the 2 situations (0.675 vs. -0.079). The same applies to animal 3, whose EBV depends entirely on the contributions of its progeny; its 2 offspring have yield deviations of -26% and -10%, and -14% and +10% in the individual and pooled models, respectively, thus explaining the difference observed in its EBV in the 2 situations (-0.7083 vs. -0.0891). The reduced variation of pooled observations not only affects the yield deviations, but also the parent averages and the progeny contributions.

The loss of information and the reduced variation in the case of pooled observations as compared with individual observations lead to decreased accuracy of the estimates, especially for individuals with phenotypic observations only (and no offspring).

Data Analysis

Variance components and breeding values for BW and egg production were estimated from the MME with a REML procedure. The Asreml software package was used for the analysis (Gilmour et al., 2002). The function *and()* in Asreml was used to fit multiple genetic effects per observation. For the analysis of individual and pooled data, respectively, the models described in Eq. [1] and [2] of the previous section were used. Because all hens were kept in a single stable and on the same tier of the battery system, the only fixed effect considered in this study was the effect of line.

RESULTS

Variance Components

Basic statistics of the data are summarized in Table 4. The number of hens in the analysis gradually de-

Table 3. Estimated breeding values and accuracies for the animals of the numerical example when analyzed individually or as pooled observations (in groups of 2)

Effect	Individual		Pooled	
	Solutions	Accuracies	Solutions	Accuracies
μ	12.1250		12.2228	
a_1	0.7083	0.3227	0.0891	0.1723
a_2	-0.4583	0.3227	-0.4257	0.3224
a_3	-0.7083	0.3227	-0.0891	0.1723
a_4	0.4583	0.3227	0.4257	0.3224
a_5	-0.3250	0.4387	-0.0792	0.3605
a_6	0.6750	0.4387	-0.0792	0.3605
a_7	0.0750	0.4387	-0.4703	0.2943
a_8	-1.0917	0.4463	-0.5594	0.3107
a_9	-0.3250	0.4822	0.3812	0.3690
a_{10}	1.2417	0.4463	0.4703	0.2943

Table 4. Descriptive statistics of the BW traits and for egg production, both for the individual observations and for the sum of the performances of 4 hens in 1 cage (pooled)¹

Item	BW19		BW27		BW43		BW51		EGGS
	Individual	Pooled	Individual	Pooled	Individual	Pooled	Individual	Pooled	Pooled
Number	2,244	561	2,128	532	1,932	483	1,776	444	371
Mean, g	1,405	5,621	1,693	6,773	1,844	7,375	1,874	7,497	1,113
SD, g	198	658	203	678	243	781	248	789	76.5

¹BW19 = BW at 19 wk; BW27 = BW at 27 wk; BW43 = BW at 43 wk; BW51 = BW at 51 wk; EGGS = total egg production for a 51-wk laying period.

creased over time because of mortality and the restriction that only cages with 4 hens were included. The average BW at 19 wk was 1,405 g and increased gradually to 1,874 g at 51 wk. The coefficient of variation for individual observations of BW was approximately 13%. As pooled observations are based on 4 individuals, standard deviations of pooled observations are expected to be twice that of the individual observations (i.e., for independent observations). Table 4 shows that the standard deviations for pooled observations were 3.2 to 3.3 times the standard deviation of individual observations, suggesting positive covariances among observations on cage mates.

Table 5 shows the estimates of the variance components and heritabilities, together with their standard errors, based on individual observations and based on pooled observations from 4 hens per cage. Estimated heritability of BW of laying hens at different ages was moderate to high and in the range of 0.4 to 0.6, which agrees with results from the literature (Besbes et al., 1992; Kabir et al., 2006). Heritability estimates of BW from individual and pooled observations were very similar at younger ages (BW19 and BW27), but started to differ at older ages (BW43 and BW51), although estimates were not significantly different. Theoretically, it

is expected that the residual variance estimated based on pooled observations is 4 times that estimated based on individual observations; results were in close agreement with this expectation for BW19 and BW27 (ratios of 4.73 and 4.27), but less so for BW43 and BW51 (ratios of 5.95 and 6.27). Table 5 shows that the standard error of the heritability estimates from pooled observations is about twice that when using individual observations: around 0.11 for pooled observations and around 0.05 when based on individual observations.

For egg production, only pooled observations were available, and, therefore, the comparison with the estimates from individual observations was not possible. The estimated heritability for egg production is 0.34, and this is in line with what has been reported in the literature (Wei and van der Werf, 1993; Anang et al., 2000). The standard error of the estimate is 0.13 and of the same order of magnitude as for BW based on pooled observations.

Breeding Values

Table 6 shows the mean accuracy of the estimated breeding values for all hens with an observation, and for sires and dams with at least 10 and 4 progeny,

Table 5. Variance components, h^2 estimates, and standard errors for BW traits and for egg production based on individual and pooled observations^{1,2}

Trait	Analysis	σ_a^2	σ_e^2	h^2	SE
BW19	Individual	10,872	8,574	0.56	0.05
	Pooled	10,747	40,618	0.51	0.11
BW27	Individual	11,059	8,052	0.58	0.05
	Pooled	10,901	34,408	0.56	0.12
BW43	Individual	16,777	15,682	0.52	0.06
	Pooled	12,219	93,289	0.34	0.10
BW51	Individual	21,449	14,488	0.60	0.06
	Pooled	16,247	90,896	0.42	0.12
EGGS	Pooled	241.32	1,878.5	0.34	0.13

¹BW19 = BW at 19 wk; BW27 = BW at 27 wk; BW43 = BW at 43 wk; BW51 = BW at 51 wk; EGGS = total egg production for a 51-wk laying period.

²Heritability estimates for pooled observations were calculated as $\frac{\sigma_A^2}{\sigma_A^2 + \frac{\sigma_e^2}{4}}$, where 4 is the number of hens per cage.

Table 6. Correlations between EBV based on individual observations and based on the pooled cage performance of 4 hens and the accuracies of the EBV for all hens with observations, sires with more than 10 offspring, and dams with more than 4 offspring

Trait ¹	Analysis	Correlations			Accuracy		
		Pooled			Mean		
		All hens with observations	Sires ≥10 offspring (96) ²	Dams ≥4 offspring (302) ²	All hens with observations	Sires ≥10 offspring	Dams ≥4 offspring
BW19	Individual	0.748	0.891	0.875	0.765	0.751	0.617
	Pooled				0.552	0.612	0.513
BW27	Individual	0.744	0.887	0.880	0.753	0.738	0.593
	Pooled				0.527	0.621	0.523
BW43	Individual	0.732	0.812	0.868	0.718	0.724	0.575
	Pooled				0.433	0.533	0.440
BW51	Individual	0.703	0.814	0.847	0.759	0.756	0.628
	Pooled				0.461	0.550	0.460
EGGS	Pooled				0.417	0.504	0.450

¹BW19 = BW at 19 wk; BW27 = BW at 27 wk; BW43 = BW at 43 wk; BW51 = BW at 51 wk; EGGS = total egg production for a 51-wk laying period.

²The number of individuals on which these correlations were based differed between traits and varied between 86 and 96 for sires and 235 and 302 for dams.

respectively. The accuracy of EBV from pooled observations is consistently less than that of EBV from individual records. The decrease in accuracy for pooled observations is greater for hens with observations than for sires with at least 10 offspring or for dams with at least 4 offspring.

Selection index theory predicts that, in the case of unrelated group mates, the accuracy of EBV based on the cage performance of n individuals (pooled data) decreases with the number of individuals per cage as

$$r_{IH} = \sqrt{\frac{h^2}{n}}, \text{ where } n \text{ is the number of unrelated cage-}$$

mates. If cages consist of half sibs and full sibs, the accuracy of EBV predicted from pooled data will asymptotically reach a value of 0.50 and 0.707 in the cases of half-sib and full-sib cage mates, respectively. To verify the decrease in accuracy when increasing the group size, we randomly pooled BW19 observations of 2, 3, 4, 6, and 8 animals originating from any cage, and calculated mean accuracies. The mean accuracies for hens with an observation were 0.766 when we used individual observations, 0.684 when based on pooled data of 2 hens, 0.569 when based on pooled data of 3 hens, 0.552 when based on pooled data of 4 hens, 0.405 when based on pooled data of 5 hens, and 0.330 when based on pooled data of 8 hens. These results show that accuracies clearly decrease if the group size increases, as was also shown by Olson et al. (2006), and as would have been expected based on selection index calculations. However, the accuracies thus calculated decrease following a different pattern than that theoretically expected on the basis of unrelated cage mates. The reason is that in the experimental data set used in this study there were not only unrelated animals, but also cages of half or full sibs; moreover, additional pedigree relationships were present.

Table 6 also shows correlations between the estimated breeding values based on individual and pooled observations for 3 groups of individuals: all hens with an observation, sires with more than 10 daughters, and dams with more than 4 daughters. The correlations between EBV from individual and pooled observations are between 0.70 and 0.75 when all hens with an observation are considered. For sires with at least 10 daughters, correlations are between 0.81 and 0.89, depending on the trait. For dams with 4 daughters, correlations between EBV were between 0.85 and 0.88.

DISCUSSION

Results show that estimation of variance components and prediction of breeding values from pooled data instead of individual observations is theoretically and practically feasible. Earlier attempts to incorporate group observations into genetic evaluations were made by Simianer and Gjerde (1991), who used the mean performance from groups of full-sibs to estimate variance components for BW in salmon, and by Nurgartining-sih et al. (2002) and Wei and van der Werf (1995), who both used cage means in a sire model for the estimation of genetic parameters for egg production traits in laying hens. Following the work of Olson et al. (2006), based on simulated data, we developed a method based on modified MME using an animal model and applied it to real data. The data available in the present study made it possible to compare genetic evaluations based on individual and pooled observations.

The results show that for the same number of animals, EBV and heritability estimates based on pooled data are less accurate than those based on individual observations. However, in practice comparison of esti-

mates based on an equal number of phenotypes rather than based on an equal number of individuals might be more appropriate. The use of pooled data offers some advantages: they are often easier and cheaper to collect, they may be in some cases the only data available, and they may sometimes better reflect the commercial environment where animals are kept, thus avoiding bias due to possible genotype $\times$ environment interaction.

At the commercial level only pooled data are available. These data may be of great interest for breeding companies to select pure line individuals. Because animals at the commercial level are almost always crossbred individuals, this aspect should be accounted for (e.g., Wei and van der Werf, 1995).

Presence of G $\times$ E interaction for animals kept individually or in groups will reduce the response to selection at the commercial level. Merks (1989) found a genetic correlation of 0.64 for backfat thickness measured on individually housed pigs at central test stations and on group-housed pigs on commercial fattening farms. In broilers, Zerehdaran et al. (2005) estimated a genetic correlation of 0.80 between BW in group housing and individual cages. These estimates clearly suggest the presence of genotype $\times$ environment interaction for individually and group-housed animals.

Therefore, selecting individuals based on group performance may be convenient in many situations. It will make genetic evaluation of farm animals more flexible and more accurate by including information collected under commercial conditions where animals are kept in groups.

Accuracies of Selection

The accuracy of selection based on pooled observations, for all hens with records and sires and dams with at least 10 or 4 offspring is, respectively, 65 to 72% and 72 to 88% of the accuracies that can be obtained when using individual observations. However, this reduced accuracy should be interpreted in the context of direct vs. indirect selection. The breeding goal of a breeding company is the trait under commercial conditions (i.e., group housing). If testing is under individual housing, the genetic correlation between group and individual housing is relevant. 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 the traits (Falconer, 1989). Similarly, the ratio between accuracies based on pooled and individual data provides a threshold for the genetic correlation between individual and group housing below which pooled data would result in a greater selection response (for the same selection intensity). Therefore, in the present study a genetic correlation below 0.72 for sires would compensate for the loss of accuracy due to pooled observations.

Fixed Effects

The relationship matrix enables estimation of the breeding value of each animal based on pooled observations. However, some differences in the definition of systematic environmental effects may arise when using either individual or pooled records. In the numerical example presented in this paper, only the fixed effect of the mean was considered and the vector of fixed effects was the same in both models, but this will not always be the case. Examples are when pooled phenotypes are being produced by animals of different sex or breeds. In these situations, the design matrix for the fixed effects must be modified to reflect the contribution of the fixed effect classes to the pooled phenotype. A consequence will be that systematic environmental effects can be estimated less accurately for pooled data when compared with a situation where individual observations are available.

Group Composition

In the present study about one-half of the cages were composed of full-sibs and the other one-half had a random composition. The numerical example used to outline the procedure illustrated that 2 full-sibs in group A (animals 5 and 6) had the same EBV when estimated based on pooled data, and, therefore, the composition of the groups will have an impact on breeding value estimation and estimation of variance components. Group observations are in fact disentangled through the pedigree structure. Family groups seem to be better for productivity and social welfare (P. Bijma, Wageningen University, Wageningen, the Netherlands, personal communication); therefore, on one hand having groups of closely related animals poses statistical challenges; on the other hand, they offer a better social environment and genetic background for animal performance and welfare. The appropriate cage composition, therefore, needs to be chosen to balance these 2 contrasting aspects.

Effects of Selection and Competition

According to theory, the residual variance estimated from pooled observations should be 4 times that estimated from individual observations. Results are in close agreement with this hypothesis for BW19 and BW27, but less so for BW43 and BW51. The reason why results diverged from theoretical expectations later in life might in part be that we limited this study to groups with 4 hens per cage. Excluding cages with less than 4 individuals might introduce a culling bias. This may be partly accounted for by using a multivariate approach, where the unselected trait (BW19) is analyzed simultaneously with the selected traits (BW27, BW43, and BW51), and all animals have observations on at least the trait on which selection is based (Pollak et al., 1984).

An alternative explanation might be the presence of within-group competition effects (Bijma et al., 2007a,b), which have been shown to play a role in several traits of agricultural interest as, for instance, survival in laying hens and growth rate (Ellen et al., 2007). The phenotype of an individual might depend not only on its own genotype, but also on the phenotypes or the genotypes of group mates and this alters the variance structure of the model. The consequences of these competition effects might accumulate over time and in this way bias the estimated genetic parameters. If competition effects play a role, this will affect the analyses based on individual observations, as well as those based on pooled data, but they might be affected differently. Therefore, in such cases, inclusion of the associative genetic effects in the analysis might be required (Bijma et al., 2007a,b).

This paper indicates that pooled observations can be used to estimate variance components and breeding values. The use of pooled data, therefore, offers interesting possibilities for breeding companies. In this study we focused on the use of pooled data from groups of the same size; there is scope for extension to groups of any size. We found that heritability estimates based on pooled data and individual observations tended to differ at older ages, which might be due to effects of selection, competition, or both.

LITERATURE CITED

- Anang, A., N. Mielenz, and L. Schöler. 2000. Genetic and phenotypic parameters for monthly egg production in White Leghorn hens. *J. Anim. Breed. Genet.* 117:407–415.
- Besbes, B., and V. Ducrocq. 2003. Pages 139–140 in *Poultry Genetics, Breeding and Biotechnology*. W. M. Muir and S. E. Aggrey, ed. CAB Int., Wallingford, Oxon, UK.
- Besbes, B., V. Ducrocq, J. L. Foulley, M. Protais, A. Tavernier, M. Tixier-Boichard, and C. Beaumont. 1992. Estimation of genetic parameters of egg production traits of laying hens by restricted maximum likelihood applied to a multiple-trait reduced animal model. *Genet. Sel. Evol.* 24:539–552.
- Bijma, P., W. M. Muir, E. D. Ellen, J. B. Wolf, and J. A. M. Van Arendonk. 2007a. Multilevel selection 2: Estimating the genetic parameters determining inheritance and response to selection. *Genetics* 175:289–299.
- Bijma, P., W. M. Muir, and J. A. M. Van Arendonk. 2007b. Multilevel selection 1: Quantitative genetics of inheritance and response to selection. *Genetics* 175:277–288.
- Box, G. E. P., and D. R. Cox. 1964. An analysis of transformations. *J. R. Stat. Soc. [Ser. A]* B26:211–243.
- Ellen, E. D., W. M. Muir, F. Teuscher, and P. Bijma. 2007. Genetic improvement of traits affected by interactions among individuals: Sib selection schemes. *Genetics* 176:489–499.
- Falconer, D. S. 1989. Pages 318–319 in *Introduction to Quantitative Genetics*. 3rd ed. Longman Scientific & Technical, New York, NY.
- Gilmour, A. R., B. J. Gogel, B. R. Cullis, S. J. Welham, and R. Thompson. 2002. ASReml User Guide Release 1.0. VSN International Ltd., Hemel Hempstead, UK.
- Henderson, C. R. 1975. Best linear unbiased estimation and prediction under a selection model. *Biometrics* 31:423–447.
- Henderson, C. R. 1984. *Applications of Linear Models in Animal Breeding*. Univ. Guelph Press, Guelph, Ontario, Canada.
- Kabir, M., O. O. Oni, G. N. Akpa, and I. A. Adeyinka. 2006. Heritability estimates and interrelationships of body weight and shank lengths in Rhode Island Red and White chickens. *Pak. J. Biol. Sci.* 9:2892–2896.
- Merks, J. W. M. 1989. Genotype $\times$ Environment interactions in pig breeding programmes. VI. Genetic relations between performances in central test, on-farm test and commercial fattening. *Livest. Prod. Sci.* 22:325–339.
- Mrode, R. A. 2005. Pages 45–48 in *Linear Models for the Prediction of Animal Breeding Values*. 2nd ed. CAB Int., Wallingford, Oxon, UK.
- Nurgiantiningsih, V., N. Mielenz, R. Preisinger, M. Schmutz, and L. Schöler. 2002. Genetic parameters for egg production and egg weight of laying hens housed in single and group cages. *Arch. Tierz.* 45:501–508.
- Olson, K. M., D. J. Garrick, and R. M. Enns. 2006. Predicting breeding values and accuracies from group in comparison to individual observations. *J. Anim. Sci.* 84:88–92.
- Pollak, E. J., J. H. J. van der Werf, and R. L. Quaas. 1984. Selection bias and multiple trait evaluation. *J. Dairy Sci.* 67:1590–1595.
- Simianer, H., and B. Gjerde. 1991. Estimating variance components from fullsib groups means. *J. Anim. Breed. Genet.* 108:270–279.
- Wei, M., and J. H. J. van der Werf. 1993. Animal model estimation of additive and dominance variances in egg production traits of poultry. *J. Anim. Sci.* 71:57–65.
- Wei, M., and J. H. J. van der Werf. 1995. Genetic correlation and heritabilities for purebred and crossbred performance in poultry egg production traits. *J. Anim. Sci.* 73:2220–2226.
- Zerehdaran, S., A. L. J. Vereijken, J. A. M. van Arendonk, and E. H. van der Waaij. 2005. Effect of age and housing system on genetic parameters for broiler carcass traits. *Poult. Sci.* 84:833–838.