Animal 20 (2026) 101872

Contents lists available at ScienceDirect

Animal
The international journal of animal biosciences

The impact of the low number of records when selecting for uniformity: a simulation study of birth weight in guinea pigs

Y.C. Rojas [a], N. Formoso-Rafferty [b,*], I. Cervantes [c], R. Ribeiro [a], J.P. Gutiérrez [c]

[a] Department of Statistics, Instituto de Ciências Exatas e Tecnológicas (ICET), Universidade Federal de Lavras (UFLA), 37203-202 Lavras MG, Brazil
[b] Department of Agricultural Production, Escuela Técnica Superior de Ingeniería Agronómica, Alimentaria y de Biosistemas, Universidad Politécnica de Madrid, Campus Ciudad Universitaria, Avda. Puerta de Hierro, n° 2-4, 28040 Madrid, Spain
[c] Department of Animal Production, Veterinary Faculty, Universidad Complutense de Madrid, Avda. Puerta de Hierro s/n, 28040 Madrid, Spain

ARTICLE INFO

Article history:
Received 21 January 2026
Revised 27 May 2026
Accepted 28 May 2026
Available online 4 June 2026

Keywords:
Data structure
Heteroscedastic models
Limited information
Robustness
Statistical approaches

ABSTRACT

Selection for uniformity in livestock species is desirable because it is associated with welfare and robustness. To address variability, it is essential to have more than one record per animal; therefore, the number of records per individual is crucial. The guinea pig, characterised by its low litter size, is a species of notable economic and nutritional importance, especially in the Andean region of South America. The objective of this study was to evaluate, through simulation, the impact of low record numbers on the performance of heteroscedastic models used in selection for uniformity, using guinea pig birth weight as an example. Different data structures that modulated litter size (**LS**), number of litters (**NL**), number of breeding animals (**BAs**), and number of generations of records were evaluated. The analysis of simulated data was performed under Frequentist and Bayesian statistical approaches. In general, the model was able to adequately estimate the variance components and predict genetic values, although its performance depended on the data structure. Under the Frequentist approach, the litter variance affecting the residual variance was biased downward in LS = 2 and LS = 3 (between −69 and −96%), although the bias decreased in LS = 5 and LS = 7 (between −10 and −44%). The BAs did not influence the bias. At NL = 2, the bias ranged from −68 to −76%, and at NL = 6, from −35 to −69%. The Bayesian method failed to converge in the scenario with two generations and NL = 2. The accuracy of the genetic values for both the trait and the residual variance was identical regardless of the statistical approach used. An increase in BAs did not lead to favourable changes in accuracy, but higher LS and NL tended to improve the accuracy of the genetic values of the trait and the residual variance. We conclude that selection for uniformity under a low number of repeated records, such as the limited number of within-litter birth weight data in guinea pigs, is feasible. However, it is recommended to carefully monitor the data structure to predict the expected response and, if possible, to use populations with the highest possible litter sizes.

Implications

Selection for uniformity might be unfeasible under a low number of repeated records, such as the case of birth weight of guinea pigs with a low litter size. The study was conducted to check it and concluded it was less efficient but possible. The number of breeding animals did not influence the selection efficiency. Predicted breeding values accuracy and selection success would increase with higher litter sizes and a higher number of litters. Accuracies were identical regardless of the Bayesian or Frequentist approach. Selection for uniformity in guinea pigs would be encouraging.

Introduction

The guinea pig is a rodent native mainly to some South American countries, but with a cosmopolitan distribution that makes it nowadays widely known. This species is particularly prized for its ease of handling and short reproductive cycle, making it a valuable resource for meat production (Sánchez-Macías et al., 2018). Peru, Ecuador, Bolivia, and Colombia are the largest producers and consumers of guinea pig meat (Donoso et al., 2025). In 2021, it was estimated that Peru produced just over 25 000 000 guinea pigs, with an annual consumption of close to 22 000 kg (Ministerio de Desarrollo Agrario y Riego, 2023).

Although guinea pigs have been widely used as laboratory animals (Padilla-Carlin et al., 2008) and a large amount of information has been generated on their physiological and reproductive charac-

* Corresponding author.
E-mail address: nora.formosorafferty@upm.es (N. Formoso-Rafferty).

https://doi.org/10.1016/j.animal.2026.101872
1751-7311/

teristics, management, and care, this information has limited usefulness in dealing with the practical and economic challenges associated with meat production in family and farm breeding systems. In this context, selection for uniformity, which aims to reduce variability in traits, has become a relevant objective of animal breeding programmes today. Uniformity has been associated with lower management and production costs in the livestock industry (Hennessy, 2005). It has also been associated with reduced mortality in prolific species and the robustness and resilience of animals to environmental challenges (Garreau et al., 2008; Blasco et al., 2017; Formoso-Rafferty et al., 2023).

To address the genetic control of residual variance, an appropriate statistical model capable of exploring it is needed. Mixed linear models used in genetic evaluation often assume that residual variance is homogeneous. However, the assumption of homogeneity does not allow the effects that affect variability to be unravelled. Therefore, models of genetically unstructured heterogeneous variances were developed, and genetic evaluation was improved by considering the heterogeneity of variances (Gianola et al., 1992; San Cristobal et al., 1993; Foulley and Quaas, 1995). These models were extended under the hypothesis that, in addition to environmental factors, there are also genetic factors that can modify residual variance (Sancristobal-Gaudy et al., 1998; Sorensen and Waagepetersen, 2003). Therefore, adjusting these models allows us to estimate not only genetic effects on the mean of the trait but also on the residual variance, which opens up the possibility of selecting for uniformity. Although the use of these models provided favourable results in species with relatively large litters (Formoso-Rafferty et al., 2016; Blasco et al., 2017; Sell-Kubiak et al., 2025), their performance in the variability of birth weight in species with low litter size, such as guinea pigs, has not yet been evaluated. Because dealing with variability implies repeated records, and only one birth weight is available for each animal, variability must be addressed by assigning the birth weight to the dam. Under these conditions, the low number of observations per dam could lead to less accurate estimates of the variance. This could compromise the estimation of variance components (Felleki et al., 2012; Pun et al., 2013), the accuracy of additive genetic values that affect residual variance, and consequently selection decisions for uniformity.

Several statistical models based on Frequentist and Bayesian approaches have been proposed to modelling the residual variance partially controlled by a genetic component. Currently, among the most widely used models are Double Hierarchical Generalised Linear Models (Rönnegård et al., 2010; Felleki et al., 2012), which are based on the theory of Lee and Nelder (2006), and Bayesian hierarchical models (Sorensen and Waagepetersen 2003). However, although there are several studies that use either the Frequentist approach or the Bayesian approach alone, there are very few studies that cover both approaches, whether with simulated or real data.

The objective of this study was to investigate, through simulation, the impact of low record numbers on the performance of heteroscedastic (**HE**) models used in selection for uniformity, using guinea pig birth weight as an example. It also evaluated how data structure can help improve this performance. Variance components and genetic values for the simulated data were estimated under Frequentist and Bayesian statistical approaches.

Material and methods

Simulated phenotypes

The data used in this research were simulated considering the HE model proposed by SanCristobal-Gaudy et al. (1998). Under this model, the residual variance is heterogeneous and partially under genetic control. The following HE model was used:

$$y_{ijk} = \mu + u_i + c_j + exp\left(\tfrac{1}{2}\left(\mu^* + u_i^* + c_j^*\right)\right)\varepsilon_{ijk}$$

where y_{ijk} is the i-th birth weight of a given individual k born from the mother i under its litter j; * refers to the parameters associated with the residual variance; μ and μ^* are the centrality measures, u_i and u_i^* are the additive genetic values of the mother; c_j and c_j^* are the litter effects; ε_{ijk} is a non-scaled residual, which is assumed to be $\varepsilon_{ijk} \sim N(0,1)$. Under this model, the vectors corresponding to additive genetic effects and litter effects will be assumed to follow a normal distribution given by:

$$\begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{u}^* \end{pmatrix} \sim N\left(\begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{0} \end{pmatrix}, \begin{pmatrix} \sigma_u^2 & \rho\sigma_u\sigma_{u^*} \\ \rho\sigma_u\sigma_{u^*} & \sigma_{u^*}^2 \end{pmatrix} \otimes \boldsymbol{A}\right),$$

$$\begin{pmatrix} \boldsymbol{c} \\ \boldsymbol{c}^* \end{pmatrix} \sim N\left(\begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{0} \end{pmatrix}, \begin{pmatrix} \sigma_c^2 & 0 \\ 0 & \sigma_{c^*}^2 \end{pmatrix} \otimes \boldsymbol{I}\right),$$

where A is the additive genetic relationship matrix and I is an identity matrix; σ_u^2 and $\sigma_{u^*}^2$ are the additive genetic variances affecting the trait and its residual variance, respectively; σ_c^2 and $\sigma_{c^*}^2$ are the litter variances affecting the trait and its residual variance, respectively; ρ is the correlation coefficient between the additive genetic effects affecting the mean and the residual variance; and $\otimes$ is the Kronecker product operation. In addition to μ and μ^*, other systematic effects were not simulated to avoid additional sources of statistical noise.

The additive genetic values (u_i and u_i^*) of the founder individuals were simulated from a normal distribution $N(0,1)$ and scaled by their corresponding genetic SDs as specified below. For subsequent generations, the additive genetic values for individuals with known parents were simulated as the average of the additive genetic values of the sire and dam plus a random Mendelian sampling term. The Mendelian sampling value was also simulated from a normal distribution $N(0,1)$ and scaled by half its genetic SD (Rönningen, 1974).

An initial reference scenario was defined (Table 1), in which the data for the founder generation consisted of 25 sires and 100 dams, unrelated, unselected, and non-inbred. These individuals were then mated to produce the first generation, with each sire mated with four different dams, as is customary in guinea pig breeding. From this last generation, 25 sires and 100 dams were sampled and mated randomly to produce the second generation, and so on, until the fourth generation was reached. The number of litters per mother was four, and the litter size was simulated from a truncated Poisson probability distribution with a mean of three and a maximum equal to the mean plus four, resulting in seven for this reference scenario. The sex of the offspring was also randomly determined. With the exception of discrete generation simulation instead of the usual overlapping, the structure assumed for litter size and number of breeding animals was based on characteristics habitually observed in the breeding of real guinea pig populations.

Different alternative simulation scenarios were defined to compare them with the reference scenario. These scenarios involved

Table 1
Simulated parameter values for the reference scenario of guinea pigs and alternative scenarios.

Parameter	Reference	Alternative
Number of generations	4	2, 6
Litter size	3	2, 5, 7
Number of breeding animals	25–100	50–200, 100–400
Number of litters per dam	4	2, 6

modulating the number of generations, litter size, number of breeding animals, and number of litters per dam (Table 1). In a previous analysis of data from a real population, it was not possible to estimate the litter components, likely due to the limited average number of records within-litter, which was around three. Therefore, the variance components estimated in the real data under a model fit that did not account for the litter effect for both the mean and the residual variance were partitioned into two identical components, and the simulation structure was improved in some scenarios to find out if this could help. Thus, for all scenarios, $\sigma^2_u = 145$, $\sigma^2_{u^*} = 0.20$, $\sigma^2_c = 145$, $\sigma^2_{c^*} = 0.20$ and $\rho = 0.25$ were used. Note that an arbitrary balanced partition was decided upon, as no other information was available on this subject. The simulated values of the variance components and genetic correlation were kept constant in all scenarios and for each defined scenario, 10 replicas were simulated.

Statistical analysis

The simulated data were analysed by fitting the same model described above, and in all cases, the simulated phenotypes were attributed to the dam. The direct genetic effect was not taken into account in order to comply with the restrictions of Bayesian software, and so it is assumed that this genetic effect is absorbed in the residual of the model, both in the simulation and the estimation processes. Additive genetic values were predicted, and variance components were estimated using Frequentist and Bayesian approaches. The estimation under the Frequentist approach was performed by fitting the following double hierarchical generalised linear model (Felleki et al., 2012):

$$\begin{pmatrix} y \\ \psi \end{pmatrix} = \begin{pmatrix} \mu \\ \mu_v \end{pmatrix} + \begin{pmatrix} Z & 0 \\ 0 & Z_v \end{pmatrix} \begin{pmatrix} u \\ u_v \end{pmatrix} + \begin{pmatrix} U & 0 \\ 0 & U_v \end{pmatrix} \begin{pmatrix} c \\ c_v \end{pmatrix} + \begin{pmatrix} e \\ e_v \end{pmatrix},$$

where y is the observation vector of the response variable and ψ is the observation vector of the linearised response variable for the residual variance; The subscript v indicates the vectors and matrices associated with the part of the variance model; μ (μ_v) is the population mean of y (ψ); u and u_v are the vectors of additive genetic effects; c and c_v are the litter effect vectors; e and e_v are the residual effect vectors; Z (Z_v) and U (U_v) are the incidence matrices associated with the additive genetic effect and the litter effect, respectively. The random effect vectors followed a normal distribution given by:

$$\begin{pmatrix} u \\ u_v \end{pmatrix} \sim \mathrm{N}\left(0, \begin{pmatrix} \sigma^2_{\mathrm{u}} & \sigma_{\mathrm{u,u_v}} \\ \sigma_{\mathrm{u,u_v}} & \sigma^2_{\mathrm{u_v}} \end{pmatrix} \otimes \mathrm{A}\right),$$

$$\begin{pmatrix} c \\ c_v \end{pmatrix} \sim \mathrm{N}\left(0, \begin{pmatrix} \sigma^2_c & 0 \\ 0 & \sigma^2_{c_v} \end{pmatrix} \otimes \mathrm{I}\right),$$

$$\begin{pmatrix} e \\ e_v \end{pmatrix} \sim \mathrm{N}\left(0, \begin{pmatrix} \mathrm{W}^{-1}\sigma^2_e & 0 \\ 0 & \mathrm{W}_v^{-1}\sigma^2_{e_v} \end{pmatrix} \otimes \mathrm{I}\right),$$

where $\mathrm{W} = \mathrm{diag}\left(\exp\left(\hat{\psi}\right)\right)^{-1}$ and $\mathrm{W_v} = \mathrm{diag}\left(\frac{1-\mathrm{h}}{2}\right)$; h is the vector of leverages; σ^2_e and $\sigma^2_{e_v}$ are scaling variances, which are expected to be unity, because W and $\mathrm{W_v}$ contain the reciprocals of the residual variances. The values for the response variable y were the simulated phenotypes, while the values for ψ were the linearised values of transformed squared residuals of y which was defined as:

$$\psi_i = \log\left(\hat{\sigma}^2_{\mathrm{e_i}}\right) + \frac{\frac{\tilde{e}^2_{\mathrm{i}}}{1-\mathrm{h_i}} - \hat{\sigma}^2_{\mathrm{e_i}}}{\hat{\sigma}^2_{\mathrm{e_i}}},$$

where $\hat{\sigma}^2_{\mathrm{e_i}}$ is the estimate of the residual variance for observation i, taking into account the estimated fixed and random effects for ψ_{i} (Felleki et al., 2012); $\hat{\mathrm{e}}^2_{\mathrm{i}}$ is the squared residual estimate for observation i; $\mathrm{h_i}$ is the leverage for observation i, that is, the i-th element of the diagonal of the hat matrix (projection matrix).

On the other hand, the estimation under the Bayesian approach was performed by fitting the following hierarchical Bayesian model.

$$\mathrm{y_i} = \mu + \mathrm{z}_i'u + \mathrm{w}_i'c + \exp\left(\frac{1}{2}\left(\mu^* + \mathrm{z}_i'u^* + \mathrm{w}_i'c^*\right)\right)\varepsilon_{\mathrm{i}}$$

where $\mathrm{y_i}$ is the i-th observed value of the response variable; the asterisk (*) denotes the parameters affecting the residual variance; μ and μ^* are the corresponding population means; u and u^* are the vectors of additive genetic effects; c and c^* are the vectors of litter effects; z_i' and w_i' are the incidence vectors associated with the additive genetic and litter effects, respectively; ε_{i} is an unscaled residual, such that $\varepsilon_{\mathrm{i}} \sim N(0,1)$. The random vectors of genetic and litter effects followed the same distributions assumed above in the section on simulated phenotypes.

The estimation under the Frequentist approach was performed using the AsReml software (Gilmour et al., 2015), while under the Bayesian approach, the GSEVM software (Ibáñez-Escriche et al., 2010) was used. Despite the genetic correlation between litter effects has shown that it must be taken into account when modelling trait variability (Sell-Kubiak et al., 2025), this correlation was not accounted in the simulation nor in the estimation, since the software based on the Bayesian approach does not allow for modelling such a correlation; however, it has been shown to provide optimal performance in real selection experiments (Formoso-Rafferty et al., 2016). For both approaches, Pearson's correlation coefficient between the true and predicted genetic values was calculated to determine the accuracy of the prediction in each replicate, and the expected genetic response was determined to be proportional to accuracy. Averages and SEs of accuracies and variance components were calculated considering the 10 replicates. Inferences about differences between estimates were done by inspecting virtual confidence intervals of each parameter defined by the mean ± 1.96 times the corresponding SEs.

Results

The results section has been structured beginning with the evaluation of the influence of litter size (**LS**), followed by the number of breeding animals (**BAs**), and finally, the number of litters per dam (**NL**), including in the three sections the influence of considering two, four, and six generations.

Litter size

The means of the estimates of the variance components, genetic correlation, and accuracies obtained under the Frequentist and Bayesian methodologies for simulated data with a variable number of litter size within three different generations are shown in Table 2, and the corresponding SEs are shown in Supplementary Table S1. Tables 2 and S1 also include the reference scenario. In the most restrictive scenarios, characterised by an LS = 2 and two simulated generations, the estimates of the additive genetic variance affecting the trait showed a negative bias, with values less than 130, representing an underestimation of around 11%. As the number of generations increased, the estimates increased progressively, reaching values of 145.89 and 148.96 when six generations were simulated. This trend was also observed in larger litter sizes. For all litter size classes, the SEs of the estimates tended to decrease as the number of generations of records increased.

Table 2
Mean estimates of variance components, genetic correlation, and accuracies for a variable number of litter sizes and generations obtained under a Frequentist and Bayesian approach in guinea pigs' birth weight.

Simulated			σ^2_u	$\sigma^2_{u^*}$	σ^2_c	$\sigma^2_{c^*}$	ρ	$r_{(u,\widehat{u})}$	$r_{(u^*,\widehat{u}^*)}$
			145.00	0.20	145.00	0.20	0.25		
LS = 2	gen = 2	Freq	129.50	0.24	125.23	0.01	0.22	0.76	0.55
		Bayes	129.62	0.24	137.24	0.18	0.21	0.76	0.55
	gen = 4	Freq	145.19	0.22	132.55	0.01	0.27	0.79	0.63
		Bayes	143.83	0.22	136.98	0.19	0.23	0.79	0.65
	gen = 6	Freq	148.96	0.23	134.07	0.01	0.24	0.82	0.66
		Bayes	145.89	0.21	140.99	0.22	0.24	0.82	0.67
LS = 3	gen = 2	Freq	133.15	0.24	144.18	0.06	0.26	0.80	0.66
		Bayes	133.22	0.23	146.83	0.21	0.25	0.80	0.66
	gen = 4	Freq	139.49	0.22	148.86	0.05	0.21	0.82	0.69
		Bayes	139.21	0.20	156.24	0.20	0.21	0.82	0.69
	gen = 6	Freq	149.38	0.23	137.23	0.06	0.28	0.83	0.70
		Bayes	148.10	0.21	139.09	0.20	0.28	0.83	0.70
LS = 5	gen = 2	Freq	140.57	0.22	140.61	0.11	0.28	0.83	0.75
		Bayes	141.64	0.22	139.29	0.18	0.28	0.83	0.75
	gen = 4	Freq	147.22	0.20	148.17	0.15	0.26	0.85	0.75
		Bayes	148.13	0.20	145.85	0.21	0.25	0.85	0.74
	gen = 6	Freq	141.73	0.21	146.36	0.14	0.26	0.85	0.76
		Bayes	141.12	0.20	144.33	0.21	0.26	0.84	0.76
LS = 7	gen = 2	Freq	134.92	0.21	148.55	0.18	0.23	0.85	0.77
		Bayes	136.34	0.20	146.33	0.21	0.22	0.85	0.77
	gen = 4	Freq	145.72	0.20	145.98	0.18	0.24	0.85	0.78
		Bayes	145.93	0.20	143.19	0.21	0.24	0.85	0.78
	gen = 6	Freq	144.29	0.20	150.19	0.18	0.24	0.87	0.79
		Bayes	143.55	0.19	147.50	0.21	0.24	0.86	0.78

σ^2_u: genetic variance of the trait; $\sigma^2_{u^*}$: genetic variance affecting residual variance; σ^2_c: litter variance of the trait; $\sigma^2_{c^*}$ litter variance affecting residual variance; ρ: genetic correlation between the mean of the trait and its residual variance; $r_{(u,\widehat{u})}$: accuracy of genetic values for the trait; $r_{(u^*,\widehat{u}^*)}$: accuracy of genetic values for residual variance. LS: litter size; gen: generation; Freq: Frequentist; Bayes: Bayesian.

In general, the additive genetic variance affecting the residual variance was consistently estimated, with a slight overestimation for LS = 2 and LS = 3, while for LS = 5 and LS = 7, the estimates were very close to the simulated value, ranging between 0.19 and 0.22. The genetic correlation between the trait and its residual variance was close to the simulated value of 0.25 and showed estimates ranging from 0.21 to 0.28, with SEs reduced from 0.07 to 0.02. The estimates of litter variance affecting the trait were slightly underestimated (between −2 and −13%) for LS = 2, while for large litter sizes, they closely approximated the simulated value. The estimate of litter effect variance affecting residual variance showed a clear difference between the two estimation approaches. The Frequentist approach produced significantly biased downward estimates at LS = 2 and LS = 3 (between −69 and −96%), although the bias decreased at LS = 5 and LS = 7 (between −10 and −44%). The Bayesian approach produced estimates around the simulated value (between 0.18 and 0.22), even for low litter sizes and few generations. With the exception of litter effect variance affecting residual variance, all confidence intervals for the parameters obtained by both methodologies included the true simulated value.

The accuracy of additive genetic values was high for the trait and moderate for residual variance. For the trait, accuracy increased from 0.76 to 0.82 at LS = 2 to 0.84–0.87 at LS = 7. Meanwhile, the accuracy of genetic values for residual variance increased from 0.55 to 0.67 in LS = 2 to 0.77–0.79 in LS = 7, reflecting a significant influence of litter size and number of generations on this parameter. In the two-generations scenario, the expected genetic response for birth weight uniformity drops by 29% when litter size is reduced from seven to two, while in the six-generations scenario, the response drops by 15%.

Number of breeding animals

The means of the estimates of variance components, genetic correlation, and accuracies for scenarios that set the litter size at three and a variable number of BAs and generations obtained under the Frequentist and Bayesian approaches are shown in Table 3, and their corresponding SEs are shown in Supplementary Table S2. The genetic variance of the trait reached values between 147.34 and 149.96 in BAs = 100–400. The genetic variance affecting the residual variance was consistently estimated across all scenarios, with values ranging from 0.19 to 0.24. Genetic correlation estimates tend to stabilise when there are more generations of records, with SEs ranging from 0.02 to 0.07. In general, for all parameters, higher SEs were observed for BAs = 25–100.

The estimates of litter effect variance affecting the mean tended to stabilise as the number of breeding animals increased, ranging from 137.23 to 156.24 in BAs = 25–100 and from 141.33 to 145.71 in BAs = 100–400. The Frequentist approach always resulted in values slightly lower than those of the Bayesian approach in all cases. For litter effect variance affecting residual variance, the Frequentist approach again underestimated dramatically (between −59 and −75%), while the Bayesian approach provided estimates that were around the true value (0.18–0.22), regardless of the number of breeding animals and the number of generations. The accuracy of genetic values for both the mean and residual variance improved when moving from two to six generations but did not appear to change when the number of breeding animals increased.

Number of litters

Table 4 shows the means of the estimates of the variance components, genetic correlation, and accuracies for scenarios that set the litter size at three and consider a variable number of litters per dam (NL) and generations obtained under a Frequentist and Bayesian approach, and Supplementary Table S3 shows their respective SEs. For simulated data with NL = 2, the estimates of all model parameters were more unstable. In this same scenario, with two generations, the Bayesian method did not achieve con-

Table 3
Mean estimates of variance components, genetic correlation, and accuracies for a variable number of breeding animals and generations obtained under a Frequentist and Bayesian approach in guinea pigs' birth weight.

Simulated			σ^2_u	$\sigma^2_{u^*}$	σ^2_c	$\sigma^2_{c^*}$	ρ	$r_{(u,\hat{u})}$	$r_{(u^*,\hat{u}^*)}$
			145.00	0.20	145.00	0.20	0.25		
25–100	gen = 2	Freq	133.15	0.24	144.18	0.06	0.26	0.80	0.66
		Bayes	133.22	0.23	146.83	0.21	0.25	0.80	0.66
	gen = 4	Freq	139.49	0.22	148.86	0.05	0.21	0.82	0.69
		Bayes	139.21	0.20	156.24	0.20	0.21	0.82	0.69
	gen = 6	Freq	149.38	0.23	137.23	0.06	0.28	0.83	0.70
		Bayes	148.10	0.21	139.09	0.20	0.28	0.83	0.70
50–200	gen = 2	Freq	140.26	0.22	144.71	0.06	0.25	0.81	0.65
		Bayes	139.42	0.20	147.16	0.20	0.25	0.81	0.65
	gen = 4	Freq	151.64	0.21	139.73	0.08	0.22	0.82	0.69
		Bayes	151.42	0.21	141.71	0.22	0.22	0.83	0.69
	gen = 6	Freq	151.68	0.21	144.44	0.08	0.24	0.83	0.68
		Bayes	150.17	0.19	146.13	0.22	0.23	0.83	0.68
100–400	gen = 2	Freq	149.96	0.21	143.17	0.05	0.20	0.82	0.65
		Bayes	148.63	0.20	145.71	0.18	0.20	0.81	0.65
	gen = 4	Freq	148.96	0.22	141.33	0.06	0.26	0.82	0.69
		Bayes	147.34	0.21	142.99	0.20	0.26	0.82	0.70
	gen = 6	Freq	148.94	0.22	142.69	0.07	0.23	0.83	0.70
		Bayes	147.70	0.20	144.53	0.20	0.23	0.83	0.71

σ^2_u: genetic variance of the trait; $\sigma^2_{u^*}$: genetic variance affecting residual variance; σ^2_c: litter variance of the trait; $\sigma^2_{c^*}$ litter variance affecting residual variance; ρ: genetic correlation between the mean of the trait and its residual variance; $r_{(u,\hat{u})}$: accuracy of genetic values for the trait; $r_{(u^*,\hat{u}^*)}$: accuracy of genetic values for residual variance; gen: generation; Freq: Frequentist; Bayes: Bayesian.

Table 4
Mean estimates of variance components, genetic correlation, and accuracies for a variable number of litters per dam and generations obtained under a Frequentist and Bayesian approach in guinea pigs' birth weight.

Simulated			σ^2_u	$\sigma^2_{u^*}$	σ^2_c	$\sigma^2_{c^*}$	ρ	$r_{(u,\hat{u})}$	$r_{(u^*,\hat{u}^*)}$
			145.00	0.20	145.00	0.20	0.25		
NL = 2	gen = 2	Freq	128.47	0.19	132.12	0.06	0.22	0.72	0.57
		Bayes[1]	–	–	–	–	–	–	–
	gen = 4	Freq	135.70	0.20	137.98	0.05	0.31	0.74	0.61
		Bayes	134.63	0.20	140.91	0.19	0.29	0.74	0.61
	gen = 6	Freq	157.42	0.23	143.63	0.06	0.25	0.77	0.61
		Bayes	154.69	0.21	146.03	0.21	0.25	0.76	0.62
NL = 4	gen = 2	Freq	133.15	0.24	144.18	0.06	0.26	0.80	0.66
		Bayes	133.22	0.23	146.83	0.21	0.25	0.80	0.66
	gen = 4	Freq	139.49	0.22	148.86	0.05	0.21	0.82	0.69
		Bayes	139.21	0.20	156.24	0.20	0.21	0.82	0.69
	gen = 6	Freq	149.38	0.23	137.23	0.06	0.28	0.83	0.70
		Bayes	148.10	0.21	139.09	0.20	0.28	0.83	0.70
NL = 6	gen = 2	Freq	134.77	0.22	142.11	0.07	0.22	0.85	0.71
		Bayes	134.86	0.21	144.28	0.21	0.22	0.85	0.71
	gen = 4	Freq	142.23	0.24	147.96	0.06	0.22	0.86	0.75
		Bayes	142.58	0.23	149.27	0.18	0.20	0.86	0.76
	gen = 6	Freq	147.09	0.21	144.19	0.06	0.22	0.87	0.75
		Bayes	145.24	0.19	146.08	0.18	0.22	0.87	0.75

σ^2_u: genetic variance of the trait; $\sigma^2_{u^*}$: genetic variance affecting residual variance; σ^2_c: litter variance of the trait; $\sigma^2_{c^*}$ litter variance affecting residual variance; ρ: genetic correlation between the mean of the trait and its residual variance; $r_{(u,\hat{u})}$: accuracy of genetic values for the trait; $r_{(u^*,\hat{u}^*)}$: accuracy of genetic values for residual variance. NL: number of litters; gen: generation. Freq: Frequentist; Bayes: Bayesian.
[1] indicates that the Bayesian approach did not achieve convergence.

vergence. Estimates of the genetic variance of the trait ranged from 128.47 to 157.40 in NL = 2, while in NL = 6, this range was reduced to 134.77–147.09. The genetic variance affecting the residual variance was consistently estimated. The genetic correlation was slightly biased downward (between –11 and –19%), even in the scenario with NL = 6. The litter variance affecting the mean tended to stabilise as the number of litters increased. The Frequentist approach underestimated the litter variance, affecting the residual variance between –68 and –76% for NL = 2, reducing the bias between –35 and –69% for NL = 6. The accuracy of genetic values for the mean increased from a minimum value of 0.72 (NL = 2) to a maximum of 0.87 (NL = 6), and those corresponding to the residual variance increased from a minimum of 0.57 (NL = 2) to a maximum of 0.76 (NL = 6).

Discussion

In this study, we set out to evaluate, through simulation, the influence of the number of records per animal when selection focuses on uniformity, as this has been shown to benefit welfare and robustness in prolific species (Blasco et al., 2017; Formoso-Rafferty et al., 2023). This is the case with the variability of birth weight in guinea pigs. To our knowledge, there are no studies that address the genetic control of residual variance with the aim of selecting for greater uniformity in this species. First, the impact of low litter size was addressed, and then, the influence of improved data structure was studied by increasing the number of breeding animals, the number of litters per dam, and the number of generations. The estimates of variance components and

genetic correlation obtained under a Frequentist and Bayesian approach were studied because there are very few studies on uniformity that compare the two approaches (Iung et al., 2020). Furthermore, since the selection response is directly proportional to the accuracy when selection is based on predicted breeding values, and with a view to using the HE model to carry out the selection process, the accuracy of genetic value prediction has also been studied.

Comparison of statistical approaches

The results obtained show that the HE model is capable of reasonably estimating the simulated parameters, although it faces significant limitations when the number of observations is limited. The additive genetic variances associated with the trait and the residual variances were adequately estimated in most scenarios, both by the Frequentist and Bayesian approaches. This indicates that the HE model can estimate these parameters with reasonable accuracy, although stability clearly improves as the amount of information increases. On the other hand, the estimation of the litter effect variance affecting the residual variance was particularly sensitive to litter size. In all scenarios, the Frequentist approach systematically underestimated this parameter, although it tended to approach the simulated value as the litter size increased (from 0.1 in LS = 2 to 0.18 in LS = 7). This suggests that the estimation algorithm used when fitting the model based on the Frequentist approach probably needs more information to adequately estimate this parameter.

Sonesson et al. (2013) found biased estimates of variance components when analysing real BW data in salmon using the Frequentist approach. These authors suggested that this may be due to the iterative algorithm used in parameter estimation. When the number of repeated observations is too small, or even with a single record per animal, there may not be enough information to fit a random effects model, both for mean and residual variance models (Felleki et al., 2012; Mulder et al., 2013). In contrast, the Bayesian approach produced estimates of litter variance on residual variance around the simulated values in all scenarios (ranging from 0.18 to 0.22), even in those simulated with a low number of generations of records, but did not provide results in scenarios with little information, particularly when two generations of records and two litters per mother were simulated. Although the Frequentist approach had difficulty obtaining estimates of litter variance affecting residual variance close to the true value, the analysis time required by this approach was dramatically shorter than that of the Bayesian approach.

The signal and magnitude of genetic correlation are important in studies of genetic heterogeneity of residual variance, mainly to guide selection decisions when uniformity of the trait is part of the breeding programme's objectives. Both approaches produced identical estimates of genetic correlation between the mean of the trait and its residual variance. The accuracy of the genetic values for both the mean of the trait and the residual variance were identical regardless of the statistical approach used, suggesting that the choice of methodology can be made interchangeably if the concern is more about achieving uniformity than estimating parameters.

Validation vs new findings

The estimation of variance components, particularly the litter variance affecting residual variance, was underestimated under the Frequentist approach, and this estimation improved only when higher litter sizes were simulated. The genetic correlation was estimated with relative stability in all simulated scenarios, with values ranging from 0.20 to 0.31. The variation in the mean estimates of genetic correlation decreases with larger litter size, a greater number of litters per dam, and a greater number of generations, suggesting that an increase in any of these factors would produce more accurate estimates of genetic correlation. Conversely, increasing the number of breeding animals together with an increase in the number of generations does not seem to improve the model's performance in addressing this parameter. It should be remembered that in this study, a fixed value of 0.25 for genetic correlation, as obtained in a real guinea pig population, was tested. However, large differences in estimates of this parameter have been reported in the literature (Hill and Mulder, 2010; Iung et al., 2020). In real data, genetic correlations for birth weight were reported to be 0.34 in a population of mice selected for environmental variability of that trait (Formoso-Rafferty et al., 2018), 0.15–0.59 in pigs (Sell-Kubiak et al., 2025), 0.39 in sheep (Sae-Lim et al., 2018), and 0.44 in beef cattle (Fina et al., 2013).

The accuracy of the genetic values that affect residual variance is a key criterion for evaluating the efficiency of selection for uniformity when based on predicted breeding values. As expected, in all scenarios, the accuracies of the genetic values for residual variance tended to be lower (ranging from 0.55 to 0.79) than those for the mean of the trait (ranging from 0.72 to 0.87). This is because variance has more statistical noise, and a large amount of information is needed to estimate the genetic values of residual variance with high accuracy (Mulder et al., 2007; Hill and Mulder, 2010).

The results showed that the accuracy of the genetic values of the residual variance increased consistently as the litter size and the number of generations of records increased (from 0.55 in LS = 2 to 0.79 in LS = 7), which should be taken into account when designing a genetic improvement programme. In species with relatively large litter sizes, studies have shown that selection for uniformity has been successful. Formoso-Rafferty et al. (2016) conducted a successful divergent selection experiment for environmental variability in birth weight in mice, and Blasco et al. (2017) also demonstrated the possibility of selecting for environmental variability in litter size in rabbits. On the other hand, the results found in the present study showed that increasing the number of litters per mother also leads to an increase in the accuracies of the genetic values of the residual variance (from 0.57 at NL = 2 to 0.76 at NL = 6), even achieving values comparable to those observed in simulated scenarios with high litter sizes, as expected. On the contrary, accuracy did not seem to improve when the number of breeding animals increased, with maximums of 0.70 in BAs = 25–100 and 0.71 in BAs = 100–400.

Application

In simulated scenarios with LS = 3 (reference scenario), as normally found in selected guinea pig populations, the accuracy of genetic values for residual variance was moderate (ranging from 0.66 to 0.70). This indicates that the predicted genetic values for residual variance can be used to select for uniformity, but with low efficiency compared to species with higher litter size. Under these circumstances, the response to selection would be limited and slow, and would be conditioned by the accumulation of information over several generations. The situation would be more complicated in native guinea pig populations, since they normally have an average litter size of two instead of three as considered in the reference scenario. This fact reflects the impact of litter size on accurately estimating genetic values of residual variance, and therefore, it would be advisable to use selected populations with the highest possible litter sizes in order to achieve a desirable response to selection. A possible alternative to overcome the low litter size in guinea pigs and obtain an adequate response to selection for uniformity would be to increase the number of parturitions

of female guinea pigs. However, this should be studied with caution, since the increase in the number of litters also takes longer.

Besides guinea pigs, the results obtained in this study could be useful to other species of zootechnical interest that are characterised by low litter sizes or a reduced number of records per individual. A clear example would be sheep, a species in which litter size is generally low and, therefore, the information available to adequately estimate the parameters associated with residual variance could be limited. In sheep breeding programmes, particularly for traits such as birth weight or weaning weight, the limited information available per ewe poses challenges similar to those observed in simulated scenarios of low litter size. Our results suggest that, under these conditions, selection for uniformity could be compromised due to the limited accuracy of the genetic values of the residual variance. Thus, these findings provide a useful framework for evaluating the feasibility of selection for uniformity in species with limited information.

Limitation

It should be noted that the interpretation of the results found in this study should be limited to the specific set of parameters used in the simulation. Conclusions about the performance of the HE model depend on the magnitudes of the simulated components of variance and genetic correlation, and could change if the variances were distributed differently or if the genetic correlation were greater, smaller, or even negative. Furthermore, our study simulated discrete generations and did not examine the effects of selection and non-random mating. However, real data sets often cover multiple generations, which often overlap, and include selection and non-random mating. In addition, real data typically exhibit other fixed effects in addition to the overall mean. Therefore, an interesting research topic for the future would be to conduct a simulation study using a real data structure from a guinea pig population.

Conclusion

The results show that the HE model is capable of adequately estimating variance components and predicting genetic values, although its performance depends on the data structure. An increase in the number of breeding animals does not provide favourable changes in accuracy, but a higher litter size and a higher number of litters tended to improve the accuracy of the genetic values of the trait and the residual variance. The accuracies of the genetic values were identical regardless of the statistical approach used. Overall, selection for uniformity under a low number of repeated records, such as the limited number of within-litter birth weight data in guinea pigs, is feasible. However, it is recommended to carefully monitor the data structure to predict the expected response and, if possible, to use populations with the highest possible litter sizes.

Supplementary material

Supplementary Material for this article (https://doi.org/10.1016/j.animal.2026.101872) can be found at the foot of the online page, in the Appendix section.

Ethics approval

Not applicable.

Data and model availability statement

None of the data were deposited in an official repository. Data are available from the corresponding author upon request and subject to institutional approval.

Declaration of generative AI and AI-assisted technologies in the writing process

During the revision process, the authors used DeepL to improve the style and grammar of individual sentences, making the text more concise. The authors then reviewed and edited the content as needed. The authors took full responsibility for the content of the publication.

Author ORCIDs

YCR: https://orcid.org/0000-0001-7750-8038.
NFR: https://orcid.org/0000-0003-2319-3485.
IC: https://orcid.org/0000-0002-0750-6480.
RR: https://orcid.org/0000-0003-4607-4964.
JPG: https://orcid.org/0000-0003-3142-2356.

CRediT authorship contribution statement

Y.C. Rojas: Writing – review & editing, Writing – original draft, Visualisation, Formal analysis, Conceptualisation. **N. Formoso-Rafferty:** Writing – review & editing, Visualisation, Conceptualisation. **I. Cervantes:** Writing – review & editing, Visualisation, Conceptualisation. **R. Ribeiro:** Writing – review & editing, Visualisation, Conceptualisation. **J.P. Gutiérrez:** Writing – review & editing, Supervision, Methodology, Conceptualisation.

Declaration of interest

None.

Acknowledgements

None.

Financial support statement

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. This study was conducted within the framework of Project PID2023-149012OB-I00, funded by the Spanish Ministry of Science, Innovation, and Universities.

References

Blasco, A., Martínez-Álvaro, M., García, M.L., Ibáñez-Escriche, N., Argente, M.J., 2017. Selection for environmental variance of litter size in rabbits. Genetics Selection Evolution 49, 1–8. https://doi.org/10.1186/s12711-017-0323-4.

Donoso, G., Galecio, J.S., Fuentes-Quisaguano, O.G., Pairis-Garcia, M., 2025. Guinea pig meat production in South America: reviewing existing practices, welfare challenges, and opportunities. Animal Welfare 34, 1–8. https://doi.org/10.1017/awf.2025.26.

Felleki, M., Lee, D., Lee, Y., Gilmour, A.R., Rönnegård, L., 2012. Estimation of breeding values for mean and dispersion, their variance and correlation using double hierarchical generalized linear models. Genetics Research 94, 307–317. https://doi.org/10.1017/S0016672312000766.

Fina, M., Ibáñez-Escriche, N., Piedrafita, J., Casellas, J., 2013. Canalization analysis of birth weight in Bruna dels Pirineus beef cattle. Journal of Animal Science 91, 3070–3078. https://doi.org/10.2527/jas.2012-5675.

Formoso-Rafferty, N., Cervantes, I., Ibáñez-Escriche, N., Gutiérrez, J.P., 2016. Genetic control of the environmental variance for birth weight in seven generations of a divergent selection experiment in mice. Journal of Animal Breeding and Genetics 133, 227–237. https://doi.org/10.1111/jbg.12174.

Formoso-Rafferty, N., de la Flor, M., Gutiérrez, J.P., Cervantes, I., 2018. Feed and reproductive efficiency differences between divergently selected lines for birthweight environmental variability in mice. Journal of Animal Breeding and Genetics 135, 378–389. https://doi.org/10.1111/jbg.12345.

Formoso-Rafferty, N., El-Ouazizi El-Kahia, L., Arias-Álvarez, M., Gutiérrez, J.P., Cervantes, I., 2023. Embryo survival and fertility differ in lines divergently

selected for birth weight homogeneity in mice. Journal of Animal Breeding and Genetics 140, 549–557. https://doi.org/10.1111/jbg.12778.
Foulley, J.L., Quaas, R.L., 1995. Heterogeneous variances in Gaussian linear mixed models. Genetics Selection Evolution 27, 211–228. https://doi.org/10.1051/gse:19950302.
Garreau, H., Bolet, G., Larzul, C., Robert-Granié, C., Saleil, G., SanCristobal, M., Bodin, L., 2008. Results of four generations of a canalising selection for rabbit birth weight. Livestock Science 119, 55–62. https://doi.org/10.1016/j.livsci.2008.02.009.
Gianola, D., Foulley, J.L., Fernando, R.L., Henderson, C.R., Weigel, K.A., 1992. Estimation of heterogeneous variances using empirical bayes methods: theoretical considerations. Journal of Dairy Science 75, 2805–2823. https://doi.org/10.3168/jds.S0022-0302(92)78044-8.
Gilmour, A.R., Gogel, B.J., Cullis, B.R., Welham, S.J., Thompson, R., 2015. ASReml user guide release 4.1 Functional specification. VSN international Ltd., Hemel Hempstead, UK. www.vsni.co.uk.
Hennessy, D.A., 2005. Slaughterhouse rules: animal uniformity and regulating for food safety in meat packing. American Journal of Agricultural Economics 87, 600–609.
Hill, W.G., Mulder, H.A., 2010. Genetic analysis of environmental variation. Genetics Research 92, 381–395. https://doi.org/10.1017/S0016672310000546.
Ibáñez-Escriche, N., Garcia, M., Sorensen, D., 2010. GSEVM vol 2: MCMC software to analyze genetically structured environmental variance models. Journal of Animal Breeding and Genetics 127, 249–251. https://doi.org/10.1111/j.1439-0388.2009.00846.x.
Iung, L.H.S., Carvalheiro, R., Neves, H.H.R., Mulder, H.A., 2020. Genetics and genomics of uniformity and resilience in livestock and aquaculture species: a review. Journal of Animal Breeding and Genetics 137, 263–280. https://doi.org/10.1111/jbg.12454.
Lee, Y., Nelder, J.A., 2006. Double hierarchical generalized linear models. Applied Statistics 55, 139–185. https://doi.org/10.1111/j.2517-6161.1996.tb02105.x.
Ministerio de Desarrollo Agrario y Riego, 2023. Cadena productiva del cuy. Retrieved on 30 May 2025 from: https://cdn.www.gob.pe/uploads/document/file/4061856/Cadena productiva de cuy.pdf.
Mulder, H.A., Bijma, P., Hill, W.G., 2007. Prediction of breeding values and selection responses with genetic heterogeneity of environmental variance. Genetics 175, 1895–1910. https://doi.org/10.1534/genetics.106.063743.
Mulder, H.A., Rönnegård, L., Fikse, W.F., Veerkamp, R.F., Strandberg, E., 2013. Estimation of genetic variance for macro- and micro-environmental sensitivity using double hierarchical generalized linear models. Genetics Selection Evolution 45, 1–14. https://doi.org/10.1186/1297-9686-45-23.
Padilla-Carlin, D.J., McMurray, D.N., Hickey, A.J., 2008. The guinea pig as a model of infectious diseases. American Association for Laboratory Animal Science 58, 324–340.
Pun, A., Cervantes, I., Nieto, B., Salgado, C., Pérez-Cabal, M.A., Ibáñez-Escriche, N., Gutiérrez, J.P., 2013. Genetic parameters for birthweight environmental variability in mice. Journal of Animal Breeding and Genetics 130, 404–414. https://doi.org/10.1111/jbg.12021.
Rönnegård, L., Felleki, M., Fikse, F., Mulder, H.A., Strandberg, E., 2010. Genetic heterogeneity of residual variance - estimation of variance components using double hierarchical generalized linear models. Genetics Selection Evolution 42, 1–10. https://doi.org/10.1186/1297-9686-42-8.
Rönningen, K., 1974. Monte carlo simulation of statistical-biological models which are of interest in animal breeding. Acta Agriculturae Scandinavica 24, 135–142. https://doi.org/10.1080/00015127409433236.
Sae-Lim, P., Jakobsen, J.H., Mulder, H.A., 2018. Uniformity in birth weight is heritable in Norwegian White Sheep. In: Proceedings of the World Congress on Genetics Applied to Livestock Production, 11–16 February 2018, Auckland, New Zealand, p. 278.
San Cristobal, M., Foulley, J.L., Manfredi, E., 1993. Inference about multiplicative heteroskedastic components of variance in a mixed linear Gaussian model with an application to beef cattle breeding. Genetics Selection Evolution 25, 3–30. https://doi.org/10.1186/1297-9686-25-1-3.
Sánchez-Macías, D., Barba-Maggi, L., Morales-delaNuez, A., Palmay-Paredes, J., 2018. Guinea pig for meat production: a systematic review of factors affecting the production, carcass and meat quality. Meat Science 143, 165–176. https://doi.org/10.1016/j.meatsci.2018.05.004.
Sancristobal-Gaudy, M., Elsen, J.M., Bodin, L., Chevalet, C., 1998. Prediction of the response tea selection for canalisation of a continuous trait in animal breeding. Genetics Selection Evolution 30, 423–451. https://doi.org/10.1051/gse:19980502.
Sell-Kubiak, E., Kasper, C., Lepori, A., Gutiérrez, J.P., Formoso-Rafferty, N., Khayatzadeh, N., Cervantes, I., 2025. Integrating three genetic dimensions relating to piglet birth weight: direct and maternal effects on the mean and genetic control of residual variance. Animal 19, 101651. https://doi.org/10.1016/j.animal.2025.101651.
Sonesson, A.K., Ødegård, J., Rönnegård, L., 2013. Genetic heterogeneity of within-family variance of body weight in Atlantic salmon (*Salmo salar*). Genetics Selection Evolution 45, 1–8. https://doi.org/10.1186/1297-9686-45-41.
Sorensen, D., Waagepetersen, R., 2003. Normal linear models with genetically structured residual variance heterogeneity: a case study. Genetical Research 82, 207–222. https://doi.org/10.1017/S0016672303006426.