Original article

Variance components for survival of piglets at farrowing using a reduced animal model

Ian M.S. White[a*], Rainer Roehe[b], Pieter W. Knap[c],
Sue Brotherstone[a,b]

[a] Institute of Evolutionary Biology, University of Edinburgh, Ashworth Laboratories, West
Mains Road, Edinburgh EH9 3JT, UK
[b] Sustainable Livestock Systems Group, Scottish Agriculture College, Bush Estate, Penicuik,
Midlothian EH26 0PH, UK
[c] PIC International Group, Ratsteich 31, D-24837 Schleswig, Germany

Abstract – Farrowing survival is usually analysed as a trait of the sow, but this precludes estimation of any direct genetic effects associated with individual piglets. In order to estimate these effects, which are particularly important for sire lines, it is necessary to fit an animal model. However this can be computationally very demanding. We show how direct and maternal genetic effects can be estimated with a simpler analysis based on the reduced animal model and we illustrate the method using farrowing survival information on 118 193 piglets in 10 314 litters. We achieve a 30% reduction in computing time and a 70% reduction in memory use, with no important loss of accuracy. This use of the reduced animal model is not only of interest for pig breeding but also for poultry and fish breeding where large full-sib families are performance tested.

reduced animal model / piglet survival

1. INTRODUCTION

The reduced animal (RA) model of Quaas and Pollack [6] was introduced to obtain predicted breeding values without the need to set up equations for all animals. This model distinguishes between animals with progeny records (parents) and those without progeny records (non-parents). The section of $\mathbf{A}^{-1}$ pertaining to non-parents is diagonal. Equations are obtained for parents only, then predicted progeny breeding values, equal to the parent average plus an estimated Mendelian effect, are obtained by back-solving. Predicting breeding values is a less computationally demanding task than estimating variance components, and since computing power has increased, this model has lost favour

---

* Corresponding author: i.m.s.white@ed.ac.uk

and nowadays is rarely used. Many animal breeders believe that the computational benefits of the RA model are outweighed by the difficulty of setting it up.

When pedigree information is available, variance components are usually estimated by fitting the animal model [8]. In some species, animals with records form large groups (litters) with common parentage. Here we revive the RA model and show how it can be used to estimate genetic and environmental variance components in this situation. An approximate version of the model can be fitted even when the only data recorded are litter mean values of piglet survival. We demonstrate that it is possible to fit an RA model with standard software for variance component or BLUP estimation.

As an example we applied the method to piglet mortality data. Our model treated survival as a trait of the piglet. The genetic influences on the trait consisted of the genetic effect of the piglet itself (the ability of the piglet to survive) and the maternal genetic effect (the ability of the sow to provide a suitable environment for survival of her offspring).

Often data are available only in the form of litter means or totals, such as the number of stillborn piglets and number of piglets born alive. This information is generally analysed using a model which treats survival as a trait of the sow and ignores the direct genetic effect of the piglet. Assuming an average litter size of 10 piglets, such a model is attractive, since the number of animals and number of equations to be solved is reduced to around 10% of the full analysis. In addition, such a model has fewer genetic effects and is perceived to be more "robust". The drawbacks to this model are that the direct genetic effect has been ignored and breeding values for service sires are not generated.

The analysis described in the next section retains the simplicity of the conventional analysis but has none of its drawbacks.

2. MATERIALS AND METHODS

2.1. Statistical methods

We assume observations on k litters of fullsib animals, with n_i animals in the ith litter. The total number of animals with records is $N = \sum n_i$.

With the RA model, we distinguish between animals with progeny records (parents) and those with no recorded progeny (non-parents). A parent with its own performance record appears twice in the dataset, initially as a member of its birth litter and later as the parent of a litter.

The breeding value for a non-parent is represented as the average breeding value of the sire and dam plus a Mendelian deviation specific to the animal,

with variance $\frac{1}{2}\sigma_a^2(1-f)$, where f is the average inbreeding of the two parents. As an approximation, we ignore the effect of inbreeding on the Mendelian variance, which is taken to be $\frac{1}{2}\sigma_a^2$.

Let p be the number of parents in the pedigree. This number is usually a small fraction of the total number of animals (parents plus non-parents). Let $\mathbf{a}$ and $\mathbf{m}$ ($p \times 1$) be vectors of breeding values for direct and maternal genetic effects for parents. It is assumed that $\text{cov}(\mathbf{a}) = \sigma_a^2\mathbf{A}$, $\text{cov}(\mathbf{m}) = \sigma_m^2\mathbf{A}$, where $\mathbf{A}$ is the numerator relationship matrix for parents. Equation (1) below describes the RA model for direct genetic effects, with in addition maternal genetic, litter, and fixed effects:

$$y = \mathbf{Xb} + \left(\mathbf{P} + \frac{1}{2}\mathbf{QS} + \frac{1}{2}\mathbf{QD}\right)\mathbf{a} + \mathbf{Dm} + \mathbf{Lu} + \mathbf{e}. \tag{1}$$

The data vector is $\mathbf{y}$ ($N \times 1$), $\mathbf{Xb}$ are fixed effects, $\mathbf{P}$ is an $N \times p$ matrix such that $P_{ij} = 1$ if the ith animal is parent j, zero otherwise, $\mathbf{Q}$ is an $N \times N$ diagonal matrix with $Q_{ii} = 1$ when animal i is a non-parent, zero otherwise, $\mathbf{S}$ and $\mathbf{D}$ are $N \times p$ matrices identifying sires and dams, respectively, in the list of parents. The matrix $\mathbf{L}$ ($N \times k$) relates records to litters, and the litter effects $\mathbf{u}$ have variance σ_u^2. The error vector $\mathbf{e}$ has variance σ_e^2 for parents and $\sigma_e^2 + \frac{1}{2}\sigma_a^2$ for non-parents (ignoring inbreeding).

Equation (1) suggests one way to set up the RA model to deal with direct genetic effects. For parents, identify the animal but set sire and dam "missing" and for non-parents, set the animal identity missing but give the sire and dam. The matrix $\mathbf{P} + \frac{1}{2}\mathbf{QS} + \frac{1}{2}\mathbf{QD}$ is then the design matrix for animals "overlaid" by half times the design matrices for sires and dams.

The RA model can be analysed with a BLUP or variance component estimation programme of sufficient flexibility. For example, it must have the capacity to "overlay" design matrices, and treat missing values in design matrices in an appropriate way. More detail is given in the appendix.

It is interesting to explore to what extent the data can be reduced to litter means without loss of information. Inspection of equation (1) shows that simply averaging results by litters will not result in a model that can be dealt with easily. However, for the subset of results relating to non-parents, $\mathbf{P} = \mathbf{0}$, $\mathbf{Q} = \mathbf{I}$, and for these records, reduction to litter means is straightforward:

$$\bar{\mathbf{y}} = \bar{\mathbf{X}}\mathbf{b} + \frac{1}{2}\left(\bar{\mathbf{S}} + \bar{\mathbf{D}}\right)\mathbf{a} + \bar{\mathbf{D}}\mathbf{m} + \mathbf{u} + \bar{\mathbf{e}} \tag{2}$$

where $\bar{\mathbf{y}}$ is a $k \times 1$ vector of litter means for non-parents, and $\bar{\mathbf{X}}$, $\bar{\mathbf{S}}$ and $\bar{\mathbf{D}}$ are design matrices relating fixed effects, sires and dams to litters, obtained by

averaging rows of **X**, **S** and **D**. This involves no loss of information provided any fixed effects which vary within litters are simple factors, such as sex of piglet.

For a litter containing n non-parents, the variance of $\bar{\mathbf{e}}$ is $(\sigma_e^2 + \frac{1}{2}\sigma_a^2)/n$.

The information in the data comprises (i) individual records for parents, (ii) litter means, and (iii) the within-litters variance, with (ii) and (iii) calculated from records for non-parents.

When there are no parents with their own performance records, analysis is based on (ii) and (iii). Overlaying of design matrices can be avoided by regrouping the genetic terms as $\frac{1}{2}\mathbf{Sa}$ and $\mathbf{D}(\frac{1}{2}\mathbf{a} + \mathbf{m})$. Correlated effects for sire of litter ($\frac{1}{2}\mathbf{a}$) and dam of litter ($\frac{1}{2}\mathbf{a} + \mathbf{m}$) can be fitted with variances $\frac{1}{4}\sigma_a^2$ and $\frac{1}{4}\sigma_a^2 + \sigma_m^2$, and covariance $\frac{1}{4}\sigma_a^2$. The analysis required is therefore an analysis of litter means, weighted by litter size, with correlated random effects for sire and dam of litter, and the covariance constrained to equal the sire variance. This allows estimation of the direct genetic variance σ_a^2 (as 4 times the sire variance), and the maternal genetic variance σ_m^2 (as the difference between the dam and sire variances). This analysis is possible even when litter means are the only data available. If there is a large number of litters, and sufficient variability in litter sizes, it will also be possible to estimate the litter variance σ_u^2 separately from the residual variance $\sigma_e^2 + \frac{1}{2}\sigma_a^2$, although it may be poorly estimated.

If data are available for individual animals, the within-litters mean square provides a separate estimate of the residual variance, and the litter variance σ_u^2 can then be estimated much more accurately. The within-litters mean square is always available with 0/1 data, since it can be calculated from the litter means (proportions).

When there are parents with their own performance records, each is included in the weighted analysis, with weight one, separately from the rest of its birth litter. It is linked to its birth litter through its litter, sire and dam records, and its breeding value is linked to its own litters through the overlaying of animal, sire and dam design matrices described above and in the appendix. We refer to this as the exact RA analysis.

The exact analysis requires identification of parents, overlaying of design matrices, and the use of various devices described in the appendix. The simpler analysis appropriate when there are no parents with their own records can be applied to all progeny, ignoring the distinction between parents and non-parents. This analysis will often capture most of the information in the data, and might be considered as a useful approximation when the number of parents with records is a small fraction of the total number of animals, or if litter

means, such as percentage survival per litter, are the only data recorded. We refer to this as the approximate RA analysis.

2.2. Example

The method is illustrated by estimating genetic and environmental components of variance associated with farrowing survival. Survival data were collected between 2000 and 2003 from two farms and four pig lines. Pedigree information was available on the parents of the litters.

A piglet born alive was scored as 1, a stillborn or mummified piglet as 0. The trait is therefore the probability of being born alive among all piglets (including stillborn and mummified piglets). The model included a direct genetic effect for each piglet, maternal genetic and litter effects. The fixed effects included were parity of litter, farm, line and farrowing date (as a separate smoothing spline for each farm-line combination).

Since genetic and environmental variances for this trait are small, we did not consider it necessary to make any special allowances for the binary nature of the trait. The analysis is therefore on the 0/1 scale. This also facilitated comparisons with previous work.

There were observations on 118 193 piglets in 10 314 litters of 6592 sows and 844 service sires (an average 11.5 animals per litter). Amongst the 7436 parents, 3565 had their own performance records for the litter in which they were reared, and there were 19 850 animals in the parents' pedigree.

The animal model was fitted in the usual way. In addition we fitted the exact and approximate versions of the RA model, using the same or equivalent starting values. Timings and memory use were compared for the three methods. All calculations were performed using ASReml [2].

3. RESULTS

All three analyses converged without problems in between 5–10 iterations. The results for the animal model and both forms of the RA model were virtually indistinguishable. Estimates and standard errors are given in Table I.

Average farrowing survival was 88.3%. The residual variance made up 93.6% of the total phenotypic variance. Of the remaining 6.4%, 3.9% was litter variance and the remainder was approximately equally divided between direct (1%) and maternal (1.5%) genetic components (Tab. I).

Table II gives the number of equations, the set-up time, time per iteration, and memory requirement for each model. Both forms of the RA model led to

Table I. Variance component estimates and standard errors.

	Genetic		Environmental		
	Direct	Maternal	Litter	Residual	Total
Full model	0.001 04	0.001 78	0.004 29	0.095 92	0.103 03
RA (exact)	0.001 04	0.001 78	0.004 29	0.095 87	0.102 98
RA (approx)	0.001 03	0.001 78	0.004 30	0.095 89	0.103 00
SE[†]	0.000 26	0.000 23	0.000 23	0.000 43	—

† for full animal model.

Table II. Computing parameters for full and reduced animal models.

	No. of	Time (min)		Memory
Model	equations	set-up[†]	iteration	(MB)
Full	279 202	16.62	1.74	512
Reduced (exact)	53 665	2.70	1.49	128
Reduced (approx)	50 100	2.11	1.20	128

† includes formation of A-inverse.

reduced computing times compared to the full animal model. For the exact and approximate versions of the RA model, the times per iteration were reduced to about 85% and 70% of that for the full model. The corresponding figures for the set-up time, which includes calculation of the A-inverse, and setting up and reordering of the mixed model equations, were more dramatic at 16% and 13%.

Neither version of the RA model required more than 128 MB of memory. The full animal model required 512 MB.

4. DISCUSSION

The full and reduced animal models are exactly equivalent, provided that full account is taken of the effect of inbreeding [4]. This requires a separate residual variance $\sigma_e^2 + \frac{1}{2}(1 - f)\sigma_a^2$ for each litter, where f is the average inbreeding coefficient for the two parents. The average parental inbreeding in our data was $\bar{f} = 0.061$, but our calculations effectively assume that $f = 0$ for all litters. Since the results for the animal model and the RA model were extremely close, there appears to be no need to allow for this small amount of inbreeding.

The main advantage of the RA model over the full animal model is the reduction in the number of equations that have to be set up and manipulated by

the programme. There is some advantage in the reduction in the size of the A-inverse that must be calculated, but this does not result in dramatic improvements in speed or memory use, since most programs will exploit the sparseness of the large A-inverse associated with the full model. A secondary advantage of the RA model is its greater transparency, in that the sources of information for various components of variance are more obvious. This may be useful when there is difficulty in obtaining convergence or uncertainty about whether to include a particular term in the model.

The disadvantages of the RA model are associated with the difficulties of setting it up. However, this does not apply to the approximate form of the model, which will often be a satisfactory substitute for either the full animal or the exact RA models.

The problem of analysing data recorded as full-sib group means was considered by Simianer and Gjerde [7]. They proposed an analysis based on averaging the relationship matrix for all animals by blocks defined by full-sib groups, but did not consider the possibility that parents might have both progeny and their own performance records. The approach described here based on the RA model is simpler and easier to implement with standard software.

With the approximate version of this model, it is possible to estimate a direct and maternal genetic effect even when performances are recorded as litter means or percentages. This substantially reduces the cost of recording for traits such as farrowing survival, without important loss of accuracy, as shown by the present study. In addition, the direct genetic effect of farrowing survival in sire lines can be estimated from mean litter records, often the only data available. This is important for sire lines because in a crossbreeding scheme only the direct genetic effect will have influence on the survival of piglets in commercial production. The RA model is particularly useful when litter sizes are large and only a small fraction of animals are selected to create the next generation. Therefore, the RA model is not only of interest for pig breeding but even more important for poultry and fish breeding where large full sib families are performance tested. With a model which includes both direct and maternal genetic effects, the number of equations for the RA analysis is reduced, relative to the full animal model, by twice the total number of non-parents. For example, an analysis of poultry data might involve 10 000 full-sib families of average size 25. If there are on average 2 parents per litter, there is a saving of 460 000 equations, with concomitant reductions in run-time and memory requirement. Even greater reductions can be expected with typical fish data sets.

Table III. Variance components: comparisons with two previous studies.

	Present study	Knol *et al.* Dam line	Sire line	Grandinson *et al.*
No. litters	10 314	2 820	2 874	1 046[†]
No. piglets	118 193	33 717	29 200	11 016
Average survival	88	93	93	94
Genetic direct	0.0010	0.0001	0.0004	—
Genetic maternal	0.0015	0.0032	0.0016	0.002
Litter	0.0040	0.0039	0.0031	0.004
Residual	0.096	0.059	0.057	0.052

† 1st parity only.

Although the emphasis of this paper is on methodology, we now briefly discuss the results on variance components for farrowing survival.

Our results were in general agreement with previous research [3, 5]. Grandinson *et al.* [3] found that the direct effect was negligible but estimated maternal genetic and litter variances similar to those of the present study. Knol *et al.* [5] analysed sire and dam lines on different farms and found a small direct genetic effect.

Some comparisons are given in Table III. The average farrowing survival rate was smaller than that found in the other studies. This may reflect a difference in the definition of the trait. Mummified piglets were included in our calculations, but it is not clear whether they were included in the other studies. All four analyses give the same litter variance, and there is approximate agreement on the size of maternal effects. Our estimate of the direct genetic component was larger than the estimate of Knol *et al.* [5].

Residual and total phenotypic variances were larger, perhaps reflecting the different average survival rates.

Our analysis confirms the previous finding (*e.g.*, Knol *et al.* [5]) that there is a small but statistically significant direct genetic effect associated with the individual piglet.

Fitness traits in livestock are generally of low heritability and the accuracy of breeding values for these traits is often increased by including information from correlated traits of higher heritability [1, 9]. Multivariate analysis is required to estimate the parameters needed for such an analysis but for t traits in a multivariate analysis we need to estimate $\frac{1}{2}t(t+1)$ genetic and residual parameters. In addition, the number of equations for genetic effects increases linearly with the number of traits, making multivariate analysis very demanding of computing resources. The analysis described here yields the same parameters

as a full animal model but does so with a smaller requirement in computing time and memory. The advantages of this analysis are likely to increase with the number of traits, and we suggest that the reduced animal model has a useful role in estimating variance components in both a univariate and a multivariate context.

REFERENCES

[1] Brotherstone S., Veerkamp R.F., Hill W.G., Predicting breeding values for herd life of Holstein-Friesian dairy cattle from lifespan and type, Anim. Sci. 67 (1998) 405–411.

[2] Gilmour A.R., Gogel B.J., Cullis B.R., Welham S.J., Thompson R., ASReml User Guide Release 1.0, VSN International Ltd, Hemel Hempstead, HP1 1ES, UK, 2002.

[3] Grandinson K., Lund M.S., Rydhmer L., Strandberg E., Genetic parameters for the piglet mortality traits crushing, stillbirth and total mortality, and their relation to birth weight, Acta Agr. Scand. A 52 (2002) 167–173.

[4] Henderson C.R., Equivalent linear models to reduce computations, J. Dairy Sci. 68 (1985) 2267–2277.

[5] Knol E.F., Ducro B.J., van Arendonk J.A.M., van der Lende T., Direct, maternal and nurse sow genetic effects on farrowing-, pre-weaning- and total piglet survival, Livest. Prod. Sci. 73 (2002) 153–164.

[6] Quaas R.L., Pollak E.J., Mixed model methodology for farm and ranch beef cattle testing programs, J. Anim. Sci. 51 (1980) 1277–1287.

[7] Simianer H., Gjerde B., Estimating variance components from fullsib group means, J. Anim. Breed. Genet. 108 (1991) 270–279.

[8] Thompson R., Sire evaluation, Biometrics 35 (1979) 339–353.

[9] Wall E., Brotherstone S., Woolliams J.A., Banos G., Coffey M.P., Genetic evaluation of fertility using direct and correlated traits, J. Dairy Sci. 86 (2003) 4093–4102.

APPENDIX

Linking two occurrences of a parent

As an example, consider a litter of sire 1 and dam 2, comprising four animals, three of which are born alive (score = 1) and one is stillborn (score = 0). Two of the survivors (animals 3 and 4) subsequently become parents. For the full animal model, these four animals would appear in data and pedigree files as

Data				Pedigree		
Animal	Sire	Dam	Score	Animal	Sire	Dam
3	0	0	1	3	1	2
4	0	0	1	4	1	2
D_1	0	0	0	D_1	1	2
D_2	0	0	1	D_2	1	2

together with subsequent pedigree lines featuring 3 and 4 as parents. Zeros denote missing values. Sire and dam would not be included in the data file when fitting the full animal model, but are shown here as missing values for comparison with the RA model. The dummy identifiers D_1, D_2 for non-parents are exchangeable within litters. The only non-dummy identifiers appearing in the animal column of data or pedigree belong to a parent with own performance record in its birth litter.

For the RA model, the relevant rows of data and pedigree are as follows:

Data				Pedigree		
Animal	Sire	Dam	Score	Animal	Sire	Dam
3	0	0	1	3	1	2
4	0	0	1	4	1	2
0	1	2	0			
0	1	2	1			

In this case the pedigree file involves only parents. For a parent in its birth litter, animal ID is specified, sire and dam are set missing. In all other cases, animal ID is set missing, and sire and dam of the litter are specified.

The non-parent results can be reduced to litter means, in which case this litter contributes three rows to the data file:

Animal	Sire	Dam	Score	Weight
3	0	0	1	1
4	0	0	1	1
0	1	2	0.5	2

Parameter constraints

Let $\theta_1 = \sigma_a^2$, $\theta_2 = \sigma_e^2$, $\theta_3 = \sigma_e^2 + \frac{1}{2}\sigma_a^2$. All three variance parameters appear in the model, and we must impose the constraint

$$\theta_3 - \theta_2 - \frac{\theta_1}{2} = 0. \tag{3}$$

This problem does not arise when using the RA model for BLUP calculations, because in that case the variance ratio σ_a^2/σ_e^2 is fixed. ASReml allows proportionality constraints but not the more general form of constraint (3). We found various solutions to this problem.

(a) Instead of having separate residual variances for parents and non-parents, we let σ_e^2 be a residual variance for all records, and set up an extra variance component specifically for non-parents. There is a simple proportional relationship between the extra component ($\frac{1}{2}\sigma_a^2$) and the additive genetic variance (σ_a^2).

A disadvantage of this approach is that it does not reduce the number of effects estimated compared to the full animal model. Each parent has an effect, and each non-parent a Mendelian component. Nevertheless we found some reduction in run times (around 10–20%) with this approach, perhaps due to the smaller A-inverse used (parents only). An advantage of this approach is that BLUP for non-parents are easily reconstructed as the sum of the average parental BLUP and the Mendelian component.

(b) Method (a) can be modified so that $\sigma_e^2 + \frac{1}{2}\sigma_a^2$ is the residual variance for all records, with an "extra" variance ($-\frac{1}{2}\sigma_a^2$) specifically for parents. This is constrained to be -0.5 times the additive genetic variance, and the default positivity constraint is removed. This adds one extra equation for each parent, but the number of extra equations is generally much less than for method (a).

(c) ASReml allows a user-supplied "OWN" variance structure. It is possible to define a residual variance for non-parents ($\sigma_e^2 + \frac{1}{2}\sigma_a^2$) in terms of two parameters, σ_e^2 and σ_a^2, to be estimated separately. The first is constrained to be equal to the residual variance for parents, the second to the additive genetic variance.

(d) A Lagrange multiplier can be used for constraints such as (3). Unfortunately this method is only feasible if programmed internally. Otherwise it requires several programme runs. Some form of *ad hoc* adjustment is required if variance components stray outside the parameter space at any stage.

Write $\boldsymbol{\theta}$ for the vector $(\theta_1, \theta_2, \theta_3)$, $\mathbf{H}'\boldsymbol{\theta} = 0$ for the constraint, and denote the score vector and information matrix by $\mathbf{S}$ and $\mathbf{B}$.

With an unconstrained analysis, the update step is $\boldsymbol{\theta} \rightarrow \boldsymbol{\theta} + \boldsymbol{\delta}$, where $\boldsymbol{\delta} = \mathbf{B}^{-1}\mathbf{S}$. The constraint can be accommodated by starting with a value of $\boldsymbol{\theta}$ satisfying the constraint and modifying the update step to $\boldsymbol{\delta} = \mathbf{B}^{-1}(\mathbf{S} - \mathbf{H}\lambda)$, where

$$\begin{pmatrix} \mathbf{B} & \mathbf{H} \\ \mathbf{H}' & \mathbf{0} \end{pmatrix} \begin{pmatrix} \delta \\ \lambda \end{pmatrix} = \begin{pmatrix} \mathbf{S} \\ \mathbf{0} \end{pmatrix}.$$

The variance-covariance matrix for the constrained parameter estimate is the leading submatrix of

$$\begin{pmatrix} \mathbf{B} & \mathbf{H} \\ \mathbf{H}' & \mathbf{0} \end{pmatrix}^{-1}.$$

All computing times reported in this paper are for method (b).