

Assessment of a Poisson animal model for embryo yield in a simulated multiple ovulation-embryo transfer scheme

RJ Tempelman^{1*}, D Gianola²,

¹ *University of Wisconsin-Madison, Department of Dairy Science;*

² *University of Wisconsin-Madison, Department of Meat and Animal Science, 1675, Observatory Drive, Madison, WI 53706, USA*

(Received 16 March 1993; accepted 10 January 1994)

Summary – Estimation and prediction techniques for Poisson and linear animal models were compared in a simulation study where observations were modelled as embryo yields having a Poisson residual distribution. In a one-way model (fixed mean plus random animal effect) with genetic variance (σ_u^2) equal to 0.056 or 0.125 on a log linear scale, Poisson marginal maximum likelihood (MML) gave estimates of σ_u^2 with smaller empirical bias and mean squared error (MSE) than restricted maximum likelihood (REML) analyses of raw and log-transformed data. Likewise, estimates of residual variance (the average Poisson parameter) were poorer when the estimation was by REML. These results were anticipated as there is no appropriate variance decomposition independent of location parameters in the linear model. Predictions of random effects obtained from the mode of the joint posterior distribution of fixed and random effects under the Poisson mixed-model tended to have smaller empirical bias and MSE than best linear unbiased prediction (BLUP). Although the latter method does not take into account nonlinearity and does not make use of the assumption that the residual distribution was Poisson, predictions were essentially unbiased. After log transformation of the records, however, BLUP led to unsatisfactory predictions. When embryo yields of zero were ignored in the analysis, Poisson animal models accounting for truncation outperformed REML and BLUP. A mixed-model simulation involving one fixed factor (15 levels) and 2 random factors for 4 sets of variance components was also carried out; in this study, REML was not included in view of highly heterogeneous nature of variances generated on the observed scale. Poisson MML estimates of variance components were seemingly unbiased, suggesting that statistical information in the sample about the variances was adequate. Best linear unbiased estimation (BLUE) of fixed effects had greater empirical bias and MSE than the Poisson estimates from the joint posterior distribution, with differences between the

* Present address: Department of Experimental Statistics, Louisiana Agricultural Experiment Station, Louisiana State University Agricultural Center, Baton Rouge, LA 70803-5606 USA

2 analyses increasing with the genetic variance and with the true values of the fixed effects. Although differences in prediction of random effects between BLUP and Poisson joint modes were small, they were often significant and in favor of those obtained with the Poisson mixed model. In conclusion, if the residual distribution is Poisson, and if the relationship between the Poisson parameter and the fixed and random effects is log linear, REML and BLUE may lead to poor inferences, whereas the BLUP of breeding values is remarkably robust to the departure from linearity in terms of average bias and MSE.

Poisson distribution / embryo yield / generalized linear mixed model / variance component estimation / counts

Résumé – Évaluation d'un modèle individuel poissonnien pour le nombre d'embryons dans un schéma d'ovulation multiple et de transfert d'embryons. *Des techniques d'estimation et de prédiction pour des modèles poissonniens et linéaires ont été comparées par simulation de nombres d'embryons supposés suivre une distribution résiduelle de Poisson. Dans un modèle à un facteur (moyenne fixée et effet individuel aléatoire) avec des variances génétiques (σ_u^2) égales à 0,056 ou 0,125 sur une échelle loglinéaire, la méthode de maximisation de la vraisemblance marginale (MML) de Poisson donne des estimées de σ_u^2 ayant un biais empirique et une erreur quadratique moyenne (MSE) inférieurs à l'analyse des données brutes, ou transformées en logarithmes, par le maximum de vraisemblance restreinte (REML). De même, la variance résiduelle (le paramètre de Poisson moyen) était moins bien estimée par le REML. Ce résultat était prévisible, car il n'existe pas de décomposition appropriée de la variance indépendante des paramètres de position dans le modèle linéaire. Les prédictions des effets aléatoires obtenues à partir du mode de la distribution conjointe a posteriori des effets fixés et aléatoires sous un modèle mixte poissonnien tendent à avoir un biais empirique et une MSE inférieurs à la meilleure prédiction linéaire sans biais (BLUP). Bien que cette dernière méthode ne prenne en compte ni la non-linéarité ni l'hypothèse d'une distribution résiduelle de Poisson, les prédictions sont sans biais notable. Le BLUP appliqué après transformation logarithmique des données conduit cependant à des prédictions non satisfaisantes. Quand les valeurs nulles du nombre d'embryons sont ignorées dans l'analyse, les modèles individuels poissonniens prenant en compte la troncature donnent de meilleurs résultats que le REML et le BLUP. Une simulation de modèle mixte à un facteur fixé (15 niveaux) et 2 facteurs aléatoires pour 4 ensembles de composantes de variance a également été réalisée; dans cette étude, le REML n'était pas inclus à cause de la nature hautement hétérogène des variances générées sur l'échelle d'observation. Les estimées MML poissonniennes sont apparemment non biaisées, ce qui suggère que l'information statistique sur les variances contenue dans l'échantillon est adéquate. La meilleure estimation linéaire sans biais (BLUE) des effets fixés a un biais empirique et une MSE supérieurs aux estimées de Poisson dérivées de la distribution conjointe a posteriori, avec des différences entre les 2 analyses qui augmentent avec la variance génétique et les vraies valeurs des effets fixés. Bien que les différences soient faibles entre les effets aléatoires prédits par le BLUP et par les modes conjoints poissonniens, elles sont souvent significatives et en faveur de ces dernières. En conclusion, si la distribution résiduelle est poissonnienne, et si la relation entre le paramètre de Poisson et les effets fixés et aléatoires est loglinéaire, REML et BLUE peuvent conduire à des inférences de mauvaise qualité, alors que le BLUP des valeurs génétiques se comporte d'une manière remarquablement robuste face aux écarts à la linéarité, en termes de biais moyen et de MSE.*

distribution de Poisson / nombre d'embryons / modèle linéaire mixte généralisé / composante de variance / comptage

INTRODUCTION

Reproductive technology is important in the genetic improvement of dairy cattle. For example, multiple ovulation and embryo transfer (MOET) schemes may aid in accelerating the rate of genetic progress attained with artificial insemination and progeny testing of bulls in the past 30 years (Nicholas and Smith, 1983).

An important bottleneck of MOET technology, however, is the high variability in quantity and quality of embryos collected from superovulated donor dams (Lohuis *et al.*, 1990; Liboriussen and Christensen, 1990; Hahn, 1992; Hasler, 1992). Keller and Teepker (1990) simulated the effect of variability in number of embryos following superovulation on the effectiveness of nucleus breeding schemes and concluded that increases of up to 40% in embryo recovery rate (percentage of cows producing no transferable embryos) could more than halve female-realized selection differentials, the effect being greatest for small nucleus units. Similar results were found by Ruane (1991). In these studies, it was assumed that residual variation in embryo yields was normal, and that yield in subsequent superovulatory flushes was independent of that in a previous flush, *ie* absence of genetic or permanent environmental variation for embryo yield.

Optimizing embryo yields could be important for other reasons as well. For instance, with greater yields, the gap in genetic gain between closed and open nucleus breeding schemes could be narrowed (Meuwissen, 1991). Furthermore, because of possible antagonisms between production and reproduction, it may be necessary to use some selection intensity to maintain reproductive performance (Freeman, 1986). Also, if yield promotants, such as bovine somatotropin, are adopted, the relative economic importance of production and reproduction, with respect to genetic selection, will probably shift towards reproduction. Finally, if cytoplasmic or nonadditive genetic effects turn out to be important, it would be desirable to increase embryo yields by selection, so as to produce the appropriate family structures (Van Raden *et al.*, 1992) needed to fully exploit these effects.

Lohuis *et al.* (1990) found a zero heritability of embryo yield in dairy cattle. Using restricted maximum likelihood (REML), Hahn (1992) estimated heritabilities of 6 and 4% for number of ova/embryos recovered and number of transferable embryos recovered, respectively, in Holsteins; corresponding repeatabilities were 23 and 15%. Natural twinning ability may be closely related to superovulatory response in dairy cattle, as cow families with high twinning rates tend to have a high ovarian sensitivity to gonadotropins, such as PMSG and FSH (Morris and Day, 1986). Heritabilities of twinning rate in Israeli Holsteins were found to be 2%, using REML, and 10% employing a threshold model (Ron *et al.*, 1990).

Best linear unbiased prediction (BLUP) of breeding values, best linear unbiased estimation (BLUE) of fixed effects, and REML estimation of genetic parameters are widely used in animal breeding research. However, these methods are most appropriate when the data are normally distributed. The distribution of embryo yields is not normal, and it is unlikely that it can be rendered normal by a transformation, particularly when mean yields are low and embryo recovery failure rates are high. Analysis of discrete data with linear models, such as those employed in BLUE or REML, often results in spurious interactions which biologically do not exist (Quaas *et al.*, 1988), which, in turn, leads to non-parsimonious models.

It seems sensible to consider nonlinear forms of analysis for embryo yield. These may be computationally more intensive than BLUP and REML, but can offer more flexibility. The study of Ron *et al* (1990) suggests that nonlinear models for twinning ability may have the potential of capturing genetic variance for reproduction that would not be usable by selection otherwise. For example, threshold models have been suggested for genetic analysis of categorical traits, such as calving ease (Gianola and Foulley, 1983; Harville and Mee, 1984). In these models, gene substitutions are viewed as occurring in a underlying normal scale. However, the relationship between the outward variate (which is scored categorically, *eg*, 'easy' *versus* 'difficult' calving) and the underlying variable is nonlinear and mediated by fixed thresholds. Selection for categorical traits using predictions of breeding values obtained with nonlinear threshold models was shown by simulation to give up to 12% greater genetic gain in a single cycle of selection than that obtained with linear predictors (Meijering and Gianola, 1985). Because genetic gain is cumulative, this increase may be substantial. The use of better models could also improve (*eg*, smaller mean squared error (MSE)) estimates of differences in embryo yield between treatments.

In the context of embryo yield, an alternative to the threshold model is an analysis based on the Poisson distribution. This is considered to be more suitable for the analysis of variates where the outcome is a count that may take values between zero and infinity. A Poisson mixed-effects model has been developed by Foulley *et al* (1987). From this model, it is possible to obtain estimates of genetic parameters and predictors of breeding values.

The objective of this study was to compare the standard mixed linear model with the Poisson technique of Foulley *et al* (1987), *via* simulation, for the analysis of embryo yield in dairy cattle. Emphasis was on sampling performance of estimators of variance components (REML *versus* marginal maximum likelihood, MML, for the Poisson model), of estimators of fixed effects and of predictors of breeding values (BLUE and BLUP evaluated at average REML estimates of variance, *versus* Poisson posterior modes evaluated at the true values of variance).

AN OVERVIEW OF THE POISSON MIXED MODEL

Under Poisson sampling, the probability of observing a certain embryo yield response (y_i) on female i as a function of the vector of parameters θ can be written as:

$$\Pr(Y_i = y_i | \theta) = \frac{\lambda_i^{y_i} \exp(-\lambda_i)}{y_i!}, \quad y_i = 0, 1, 2, \dots; \lambda_i > 0 \quad [1]$$

with

$$E[Y_i | \lambda_i] = \text{Var}[Y_i | \lambda_i] = \lambda_i \quad [2]$$

The Poisson mixed model introduced by Foulley *et al* (1987) makes use of the link function of generalized linear models (McCullagh and Nelder, 1989). This function allows the modelling the Poisson parameter λ_i for female i in terms of θ . This parameterization differs from that presented in Foulley *et al* (1987) who modelled Poisson parameters for individual offspring of each female, allowing for extension to a bivariate Poisson-binomial model. The univariate Poisson model was also used

in Foulley and Im (1993) and Perez-Enciso *et al* (1993). In the Poisson model, the link is the logarithmic function.

$$\eta_i = \ln(\lambda_i), \quad -\infty < \eta_i < \infty, \quad 0 < \lambda_i < \infty \tag{3}$$

such that

$$\eta_i = \mathbf{w}'_i \boldsymbol{\theta} = [\mathbf{x}'_i \ \mathbf{z}'_i] \begin{bmatrix} \boldsymbol{\beta} \\ \mathbf{u} \end{bmatrix} \tag{4}$$

Above, $\boldsymbol{\theta}' = [\boldsymbol{\beta}', \mathbf{u}']$, $\boldsymbol{\beta}$ is a $p \times 1$ vector of fixed effects, \mathbf{u} is a $q \times 1$ vector of breeding values, and $\mathbf{w}'_i = [\mathbf{x}'_i \ \mathbf{z}'_i]$ is an incidence row vector relating $\boldsymbol{\theta}$ to η_i . Let $\mathbf{X} = \{\mathbf{x}'_i\}$ and $\mathbf{Z} = \{\mathbf{z}'_i\}$ be incidence matrices of order $n \times p$ and $n \times q$, respectively, such that:

$$\boldsymbol{\eta} = \{\eta'_i\} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} \tag{5}$$

In a Bayesian context, Foulley *et al* (1987) employ the prior densities

$$p(\boldsymbol{\beta}) \propto \text{constant} \tag{6a}$$

and

$$\mathbf{u} | \mathbf{A}, \sigma_u^2 \sim N(\mathbf{0}, \mathbf{A}\sigma_u^2) \tag{6b}$$

where \mathbf{A} is the matrix of additive relationships between animals and σ_u^2 is the additive genetic variance. Given σ_u^2 , Foulley *et al* (1987) calculate the mode of the joint posterior distribution of $\boldsymbol{\beta}$ and \mathbf{u} with the algorithm

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}\mathbf{X} & \mathbf{X}'\mathbf{R}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}\mathbf{Z} & \mathbf{Z}'\mathbf{R}\mathbf{Z} + \frac{\mathbf{A}^{-1}}{\sigma_u^2} \end{bmatrix}_{\theta=\theta^{[t-1]}} \begin{bmatrix} \boldsymbol{\beta}^{[t]} \\ \mathbf{u}^{[t]} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}\bar{\mathbf{y}} \\ \mathbf{Z}'\mathbf{R}\bar{\mathbf{y}} \end{bmatrix} \tag{7}$$

where t denotes iterate number,

$$\begin{aligned} \mathbf{R}_{n \times n} &= \text{Diag}\{\lambda_i\} \\ \bar{\mathbf{y}} &= \boldsymbol{\eta} + \mathbf{R}^{-1}[\mathbf{y} - \mathbf{E}(\mathbf{y}|\boldsymbol{\theta})] = \boldsymbol{\eta} + \left\{ \frac{y_i - \lambda_i}{\lambda_i} \right\} \end{aligned} \tag{8}$$

and where \mathbf{y} is the vector of observations. Note that the last term in [8] can be regarded as a vector of standardized (with respect to the conditional Poisson variance) residuals.

Marginal maximum likelihood (MML), a generalization of REML, has been suggested for estimating variance components in nonlinear models (Foulley *et al*, 1987; Höschele *et al*, 1987). An expectation-maximization (EM) type iterative algorithm is involved whereby

$$\sigma_u^{2[k+1]} = \frac{(\hat{\mathbf{u}}' \mathbf{A}^{-1} \hat{\mathbf{u}})^{[k]} + \mathbf{T}^{[k]}}{q} \tag{9}$$

where $T = \text{trace}(\mathbf{A}^{-1}\mathbf{C}_{\mathbf{uu}})$, such that

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}\mathbf{X} & \mathbf{X}'\mathbf{R}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}\mathbf{Z} & \mathbf{Z}'\mathbf{R}\mathbf{Z} + \frac{\mathbf{A}^{-1}}{\sigma_u^2} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{C}_{\beta\beta} & \mathbf{C}_{\beta\mathbf{u}} \\ \mathbf{C}_{\mathbf{u}\beta} & \mathbf{C}_{\mathbf{uu}} \end{bmatrix} \quad [10]$$

and $\hat{\mathbf{u}}$ is the \mathbf{u} -component solution to [7] upon convergence for a given σ_u^2 value. In [9], k pertains to the iteration number, and iterations continue until the difference between successive iterates of [9], separated by nested iterates of [7], becomes arbitrarily small. The above implementation of MML is not exact, and arises from the approximation (Foulley *et al*, 1990)

$$\mathbf{u}|\mathbf{y}, \sigma_u^2 \sim N(\hat{\mathbf{u}}, \mathbf{C}_{\mathbf{uu}}) \quad [11]$$

SIMULATION EXPERIMENTS

A one-way random effects model

Embryo yields in two MOET closed nucleus herd breeding schemes were simulated. Breeding values (\mathbf{u}) for embryo yields for n_s and n_d base population sires and dams, respectively, were drawn from the distribution $\mathbf{u} \sim N(\mathbf{0}, \mathbf{I}\sigma_u^2)$, where σ_u^2 had the values specified later. The dams were superovulated, and the number of embryos collected from each dam were independent drawings from Poisson distributions with parameters:

$$\lambda_d = \exp[\mathbf{1}\mu + \mathbf{u}_d] \quad [12]$$

where $\mathbf{1}$ is a $n_d \times 1$ vector of ones, μ is a location parameter and \mathbf{u}_d is the vector of breeding values of the n_d dams. In nucleus 1, $\mu = \ln(2)$, whereas in nucleus 2 $\mu = \ln(8)$. Note from [12] that for a given donor dam d_i ,

$$\ln(\lambda_{d_i}) = \mu + u_{d_i}, \quad i = 1, 2, \dots, n_d \quad [13a]$$

so, in view of the assumptions,

$$E(\ln(\lambda_{d_i})) = \mu \quad [13b]$$

which implies that the location parameter μ can be interpreted as the mean of the natural log of the Poisson parameters in the population of donor dams. It should be noted, as in Foulley and Im (1993) that

$$E(\lambda_{d_i}) = E(\exp(\mu + u_{d_i})) = \exp(\mu)\exp(E(u_{d_i})) = \exp(\mu)\exp\left(\frac{\sigma_u^2}{2}\right) \quad [13c]$$

Thus

$$E(\lambda_{d_i}) \geq \exp(\mu)$$

The sex of the embryos collected from the donor dams was assigned at random (50% probability of obtaining a female), and the probability of survival of a female

embryo to age at first breeding was $\pi = 0.70$ in nucleus 1, and $\pi = 0.60$ in nucleus 2. This is because research has suggested that embryo quality and yield from a single flush tend to be negatively associated (Hahn, 1992). Thus the expected number of female embryos surviving to age at first breeding produced by a given donor dam d_i is, for $i = 1, 2, \dots, n_d$,

$$E(n_i^f | \lambda_{d_i}) = \frac{\pi}{2} \lambda_{d_i}$$

and, on average,

$$\begin{aligned} E(n_i^f) &= \frac{\pi}{2} E(\lambda_d) = \frac{\pi}{2} E(\exp(\mu + u_{d_i})) = \frac{\pi}{2} \exp(\mu) \exp\left(\frac{\sigma_u^2}{2}\right) \\ &\geq \frac{\pi}{2} \exp(\mu) = \begin{cases} \pi = 0.70 & \text{for nucleus 1} \\ 4\pi = 2.40 & \text{for nucleus 2} \end{cases} \end{aligned}$$

The genetic merit for embryo yield for the i th female offspring, u_{o_i} , was generated by randomly selecting and mating sires and dams from the base population, and using the relationship:

$$u_{o_i} = \frac{1}{2} u_{s_i} + \frac{1}{2} u_{d_i} + z \sqrt{\frac{1}{2} \sigma_u^2}$$

where u_{s_i} and u_{d_i} are the breeding values of the sire and dam, respectively, of offspring i , and the third term is a Mendelian segregation residual; $z \sim N(0, 1)$. As with the dams, the vector of true Poisson parameters for female offspring was $\lambda_o = \exp[\mathbf{1}\mu + \mathbf{u}_o]$ where \mathbf{u}_o represents the vector of daughters' genetic values. The unit vector $\mathbf{1}$ in this case would have dimension equal to the number of surviving female offspring. Embryo yields for daughters were sampled from a Poisson distribution with parameter equal to the i th element of λ_o .

Four populations were simulated, and each was replicated 30 times: 1) nucleus 1 ($\mu = \ln 2$), $\sigma_u^2 = 0.056$; 2) nucleus 1 ($\mu = \ln 2$), $\sigma_u^2 = 0.125$; 3) nucleus 2 ($\mu = \ln 8$), $\sigma_u^2 = 0.056$; and 4) nucleus 2 ($\mu = \ln 8$), $\sigma_u^2 = 0.125$. Features of the 2 nucleus herds are in table I. The expected nucleus size is slightly greater than $n_s + n_d(1 + (\pi/2)\exp(\mu))$, ie about 218 cows in each of the 2 schemes, plus the corresponding number of sires. The values of σ_u^2 were arrived at as follows: Foulley *et al* (1987), using a first order approximation, introduced the parameter

$$'h^2' = \frac{\sigma_u^2}{\sigma_u^2 + E^{-1}(\lambda_i)} \quad [14]$$

which can be viewed as a 'pseudo-heritability'. Using this, the values of σ_u^2 in the 4 populations correspond to: 1) ' h^2 ' = 0.10; 2) ' h^2 ' = 0.20; 3) ' h^2 ' = 0.31; and 4) ' h^2 ' = 0.50.

In each of the 30 replicates of each population, variance components for embryo yield were estimated employing the following methods: 1) Poisson MML as in Foulley *et al* (1987); 2) REML as if the data were normal; 3) truncated Poisson MML excluding counts of zero, and using the formulae of Foulley *et al*, (1987); 4) REML-0, ie REML applied to the data excluding counts of zero; and (5) REML-LOG, which was REML applied to the data following a log transformation of the

Table I. Features of the MOET schemes simulated with a one-way random effects model.

	<i>Nucleus 1</i>	<i>Nucleus 2</i>
1. $E(\lambda_i)^*$	2	8
2. Number of base sires n_s	16	8
3. Number of base dams n_d	128	64
4. Expected embryo yield*	256	512
5. Expected number of female offspring $E(n_i^f)^{**}$	256×0.35	512×0.30
6. Expected nucleus size [†]	234	226

* Approximate, see text; ** approximate, numbers vary from replicate to replicate, see text; [†] approximate, expected nucleus size = 2 + 3 + 5.

non-null responses while discarding the null responses. Empirical bias and mean squared error (MSE) of the estimates, calculated from the 30 replicates, were used for assessing performance of the variance component estimation procedure. Because the probability of observing a zero count in a Poisson distribution with a mean of 8 is very low, the truncated Poisson and REML-0 analyses were not carried out in nucleus 2. Likewise, breeding values were predicted using the following methods: 1) the Poisson model as in [7] with the true σ_u^2 , and taking as predictors $\lambda = \exp[\mathbf{1}\hat{\mu} + \hat{\mathbf{u}}]$, where \mathbf{u} is the vector of breeding values of sires, dams, and daughters; 2) BLUP ($\mathbf{1}\mu^* + \mathbf{u}^*$) in a linear model analysis where the variance components were the average of the 30 REML estimates obtained in the replications and the asterisk denotes direct estimation of location parameters on the observed scale; 3) a truncated Poisson analysis with the true σ_u^2 and predictors as in 1); 4) BLUP-0, as in 2) but excluding zero counts, and using the average of the 30 REML-0 estimates as true variances; and 5) BLUP-LOG, as in 2) after excluding zero counts and transforming the remaining records into logs. The average of the 30 REML-LOG estimates of variance components was used in this case. BLUP-LOG predictors of breeding values were expressed as $\exp[\mathbf{1}\hat{\mu} + \hat{\mathbf{u}}]$ where $\hat{\mu}$ and $\hat{\mathbf{u}}$ are solutions to the corresponding mixed linear model equations. Hence, all 5 types of predictions were comparable because breeding values are expressed on the observed scale. As given in [12], the vector of true Poisson parameters or breeding values for all individuals was deemed to be $\lambda = \exp[\mathbf{1}\mu + \mathbf{u}]$. Average bias and MSE of prediction of breeding values of dams and daughters were computed within each data set and these statistics were averaged again over 30 further replicates. Rank correlations between different estimates of breeding values were not considered as they are often very large in spite of the fact that one model may fit the data substantially better than the other (Perez-Enciso *et al*, 1993).

A mixed model with two random effects

The base population consisted of 64 unrelated sires and 512 unrelated dams, and the genetic model was as before. The probability of a daughter surviving to age at first breeding was $\pi = 0.70$.

Embryo yields on dams and daughters were generated by drawing random numbers from Poisson distributions with parameters:

$$\lambda_{ijk} = \exp(\mu + H_i + s_j + u_k) \quad [15]$$

where μ is a fixed effect common to all observations, $\mathbf{H} = \{H_i\}$ is a 15×1 vector of fixed effects, $\mathbf{s} = \{s_j\} \sim N(\mathbf{0}, \mathbf{I}\sigma_s^2)$ is a 100×1 vector of unrelated 'service sire' effects, $\mathbf{u} = \{u_k\} \sim N(\mathbf{0}, \mathbf{A}\sigma_u^2)$ is a vector of breeding values independent of service sire effects, and σ_s^2 and σ_u^2 are appropriate variance components.

The values of $\mu + H_i$ were assigned such that:

$$E(\lambda_{ijk}) \approx \exp(\mu + H_i) = i; \quad i = 1, 2, \dots, 15 \quad [16]$$

Thus, in the absence of random effects, the expected embryo yield ranged from 1 to 15. Each of the 15 values of $\mu + H_i$ had an equal chance of being assigned to any particular record.

Service sire has been deemed to be an important source of variation for embryo yield in superovulated dairy cows (Lohuis *et al.*, 1990; Hasler, 1992). However, no sizable genetic variance has been detected when embryo yield is viewed as a trait of the donor cow (Lohuis *et al.*, 1990; Hahn, 1992). This influenced the choice of the 4 different combinations of true values for the variance components considered. In all cases, the service sire component was twice as large as the genetic component. The sets of variance components chosen were: (A) $\sigma_u^2 = 0.0125$, $\sigma_s^2 = 0.0250$; (B) $\sigma_u^2 = 0.0250$, $\sigma_s^2 = 0.0500$; (C) $\sigma_u^2 = 0.0375$, $\sigma_s^2 = 0.0750$; and (D) $\sigma_u^2 = 0.0500$, $\sigma_s^2 = 0.1000$. Along the lines of [14], the genetic variances correspond to 'pseudoheritabilities' of 7.5–22%, and to relative contributions of service sires to variance of 15–44%; these calculations are based on the approximate average true fixed effect $\bar{\lambda}$ on the observed scale in the absence of overdispersion:

$$\bar{\lambda} \approx \sum_{i=1}^{15} \exp(\mu + H_i) = 8$$

For each of the 4 sets of variance parameters, 30 replicates were generated to assess the sampling performance of Poisson MML in terms of empirical bias and square root MSE. Relative bias was empirical bias as a percentage of the true variance component. Coefficients of variation for REML and MML estimates of variance components were used to provide a direct comparison as they are expressed on different scales. REML estimates were also required in order to compare estimates of fixed effects and predictions of random effects obtained under a linear model analysis with those found under the Poisson model. MML and REML estimates were computed by Laplacian integration (Tempelman and Gianola, 1993) using a Fortran program that incorporated a sparse matrix solver, SMPAK (Eisenstat *et al.*, 1982) and ITPACK subroutines (Kincaid *et al.*, 1982) to set up the system of equations [7]. For REML, this corresponds to the derivative-free algorithm described by Graser *et al.* (1987) with a computing strategy similar to that in Boldman and Van Vleck (1991).

As in the one-way model, averages of REML estimates of the variance components obtained in 30 replicates were used in lieu of the 'true' values (which are not

well defined) to compute estimates of fixed effects and predictions of random effects in the linear model analysis; for the Poisson model, the true values of the variance components were used. Empirical biases and MSEs of the estimates of fixed effects obtained with the linear and with the Poisson models were assessed from another 30 replicates within each set of variance components. One more replicate was then generated for each variance component set, from which the empirical average bias and MSE of prediction of service sire and animal random effects were evaluated.

In order to make comparisons on the same scale, the Poisson model predictands of the random effects were defined to be $b \cdot \exp(\mathbf{s})$ for service sires and $b \cdot \exp(\mathbf{u})$ for additive genetic effects, respectively; b is the 'baseline' parameter:

$$b = \sum_{i=1}^{15} \frac{\exp(\mu + H_i)}{15} \quad [17]$$

In view of [15],

$$\lambda_{ijk} = \exp(\mu + H_i) \exp(s_j) \exp(u_k) \quad [18]$$

so that

$$E(\lambda_{ijk}) \approx \exp(\mu + H_i) \quad [19]$$

Hence,

$$E\left(\sum_{i=1}^{15} \frac{\lambda_{ijk}}{15}\right) \approx b$$

The 'baseline' value can then be interpreted as the expected value of the Poisson parameter of an observation made under the conditions of an 'average' level of the fixed effects and in the absence of random effects. The Poisson mixed-model predictions were constructed by replacing the unknown quantities in b , $\exp(\mathbf{s})$, and $\exp(\mathbf{u})$ by the appropriate solutions in [7].

In the linear mixed model, the predictors were defined to be:

$$\mathbf{1} \left(\sum_{i=1}^{15} \frac{\hat{\mu}^* + \hat{H}_i^*}{15} \right) + \hat{\mathbf{s}}^* \quad [20a]$$

and

$$\mathbf{1} \left(\sum_{i=1}^{15} \frac{\hat{\mu}^* + \hat{H}_i^*}{15} \right) + \hat{\mathbf{u}}^* \quad [20b]$$

for service sire and genetic effects, respectively. Here the unit vectors $\mathbf{1}$ are of the same dimension as the respective vectors of random effects and the asterisk is used to denote direct estimation of location parameters on the observed scale.

Estimators for fixed effects were also expressed on the observed scale. The true values of the fixed effects were deemed to be $i = \exp(\mu + H_i)$ for $i = 1, 2, \dots, 15$ as in [16]. Estimators for fixed effects under the Poisson model were therefore taken to be $\exp(\hat{\mu} + \hat{H}_i)$ for $i = 1, 2, \dots, 15$. As the linear mixed model estimates parameters on an observable scale, estimators for fixed effects were taken to be $\hat{\mu}^* + \hat{H}_i^*$.

RESULTS AND DISCUSSION

One-way model

Means and standard errors of estimates of the genetic variance (σ_u^2) for the five procedures are given in table II and MSEs of the estimates are given in table III. Clearly, estimates obtained with REML and REML-0 were extremely biased; this is so because the genetic components obtained are not on the appropriate scale of measurement (*ie* the canonical log scale). The problem was somewhat corrected by a logarithmic transformation of the records. For $E(\lambda_i) \approx 2$ and $\sigma_u^2 = 0.056$, the REML-LOG, Poisson and Poisson-truncated estimators were nearly unbiased (within the limits of Monte-Carlo variance), but the Monte-Carlo standard errors were much larger for REML-LOG. For $\sigma_u^2 = 0.125$, the Poisson estimates were biased downwards ($P < 0.05$) for both values of $E(\lambda_i)$, while those of REML-LOG were biased upwards and significantly so with $E(\lambda_i) \approx 8$. In a one-way sire threshold model, Höschele *et al* (1987) also found downward biases for the MML procedure. In spite of these small biases, however, the MSEs of the Poisson estimates (table III) were much lower than those of REML-LOG. The very large (relative to Poisson and REML-LOG) MSEs of the REML and REML-0 procedures illustrate the pitfalls incurred in carrying out a linear model analysis when the situation dictates a nonlinear analysis, or a transformation of the data.

Table II. Empirical means and standard errors of estimates (30 replicates) of genetic variance (σ_u^2) in the one-way model.

Analysis ^a	$\sigma_u^2 = 0.056$		$\sigma_u^2 = 0.125$	
	$E(\lambda_i) \approx 2$	$E(\lambda_i) \approx 8$	$E(\lambda_i) \approx 2$	$E(\lambda_i) \approx 8$
Poisson	0.047 ± 0.007	0.052 ± 0.003	0.110* ± 0.007	0.115* ± 0.004
REML	0.260** ± 0.051	3.69** ± 0.427	0.597* ± 0.065	8.40** ± 0.546
REML-LOG	0.067 ± 0.013	0.076* ± 0.010	0.136 ± 0.019	0.151* ± 0.011
Poisson-truncated	0.053 ± 0.007	na	0.103** ± 0.001	na
REML-0	0.210** ± 0.041	na	0.502** ± 0.054	na

^a See text; * different ($P < 0.05$) from true value based on *t*-test; ** different ($P < 0.01$) from true value based on *t*-test; na: not applicable.

A linear one-way random effects model, however, can be contrived in which case it can be shown that REML may actually estimate somewhat meaningful variance components on the observed scale. Presuming that multiple records on an individual is possible, the variance of Y_{ij} (with subscripts denoting the *j*th record on the *i*th individual) can be classically represented as:

$$\text{Var}(Y_{ij}) = E[\text{Var}(Y_{ij}|u_i)] + \text{Var}[E(Y_{ij}|u_i)] \tag{21a}$$

which from [2] can be written as:

$$\text{Var}(Y_{ij}) = E[\exp(\mu + u_i)] + \text{Var}[\exp(\mu + u_i)] \tag{21b}$$

Table III. Empirical (30 replicates) mean squared errors of genetic variance (σ_u^2) relative to Poisson analysis (Poisson = 1) in the one-way model.

Analysis ^a	$\sigma_u^2 = 0.056$		$\sigma_u^2 = 0.125$	
	$E(\lambda_i) \approx 2$	$E(\lambda_i) \approx 8$	$E(\lambda_i) \approx 2$	$E(\lambda_i) \approx 8$
Poisson	1.00	1.00	1.00	1.00
REML	92.3	52 857	221	130 678
REML-LOG	3.69	8.86	6.88	6.61
Poisson-truncated	1.23	na	2.13	na
REML-0	57.5	na	145	na

^a See text; na: not applicable.

such that from [13c] and results presented by Foulley and Im (1993):

$$\begin{aligned} \text{Var}(Y_{ij}) &= \exp\left(\mu + \frac{\sigma_u^2}{2}\right) + \exp(2\mu)\exp(\sigma_u^2)[\exp(\sigma_u^2) - 1] \\ &= \sigma_e^{2*} + \sigma_u^{2*} \end{aligned} \tag{21c}$$

The covariance between different records on the same individual (*ie* $\text{cov}(Y_{ij}, Y_{ij'})$) can be used to represent the variance of the random effects.

$$\text{Cov}(Y_{ij}, Y_{ij'}) = E[\text{Cov}(Y_{ij}, Y_{ij'}|u_i)] + \text{cov}[E(Y_{ij}|u_i), E(Y_{ij'}|u_i)] \tag{22a}$$

Given independent Poisson sampling conditional on u_i , the first term of the above equation is null, and

$$\begin{aligned} \text{Cov}(Y_{ij}, Y_{ij'}) &= \text{Cov}[E(Y_{ij}|u_i), E(Y_{ij'}|u_i)] \\ &= \text{Var}[\exp(\mu + u_i)] \\ &= \exp(2\mu)\exp(\sigma_u^2)[\exp(\sigma_u^2) - 1] \\ &= \sigma_u^{2*} \end{aligned} \tag{22b}$$

Thus a one-way random linear model that has the same first and second moments as Y_{ij} is

$$Y_{ij} = \mu^* + u_i^* + e_{ij}^*$$

where Y_{ij} is the j th record observed on the i th animal, μ is the overall mean, u_i^* is the random effect of the i th animal and e_{ij}^* is the residual associated with the j th record on the i th animal. Here $\mu^* = \exp(\mu + \sigma_u^2/2)$, u_i^* has null mean and variance $\sigma_u^{2*} = \exp(2\mu)\exp(\sigma_u^2)[\exp(\sigma_u^2) - 1]$ and e_{ij}^* has null mean and variance $\sigma_e^{2*} = \exp(\mu + \sigma_u^2/2)$. The empirical mean REML estimates reported in table II closely relate to the functionals for σ_u^{2*} in [22b], in spite of the violated independence assumption between genetically related random effects in the animal model.

Tables IV and V give the empirical means and MSEs, respectively, of the estimates of residual variance. It should be noted that the approximation $\exp(\mu)$

underestimated $E(\lambda_i)$, as expected theoretically. In the Poisson model, the residual variance is the Poisson parameter of the observation in question. Hence the residual variance in a linear model analysis would be comparable to $E(\lambda_i)$. The log-transformed REML estimates (REML-LOG) have no meaning here because the Poisson residual variance is generated on the observed scale, contrary to the genetic variance which arises on a logarithmic scale. Generally, the Poisson and REML methods gave seemingly unbiased estimates of the true average Poisson parameter. However, REML estimates of residual variance, rather, of $E(\lambda_i)$, appeared to be biased upwards ($P < 0.01$) for the higher genetic variance and higher Poisson mean population (table IV). The MSEs of Poisson estimates of average residual variances were much smaller than those obtained with REML, especially in the populations with a higher mean. REML-0 was even worse than REML, both in terms of bias (table IV) and MSE (table V). This is due to truncation of the distribution (eg, Carriquiry *et al.*, 1987) which is not taken into account in REML-0. The truncated Poisson analysis gave upwards biased estimates and had higher MSE than the standard Poisson method. However, truncated Poisson outperformed REML in an MSE sense in estimating the average residual variance, in spite of using less data (zero counts not included).

Table IV. Empirical means and standard errors of estimates (30 replicates) of 'average' residual variance in the one-way model.

Analysis ^a	$\sigma_u^2 = 0.056$				$\sigma_u^2 = 0.125$			
	$E(\lambda_i) \approx 2$		$E(\lambda_i) \approx 8$		$E(\lambda_i) \approx 2$		$E(\lambda_i) \approx 8$	
True average λ_i	2.084	± 0.008	8.324	± 0.056	2.180	± 0.013	8.712	± 0.073
Poisson	2.090	± 0.020	8.341	± 0.074	2.175	± 0.019	8.701	± 0.084
REML	2.030	± 0.058	8.460	± 0.277	2.130	± 0.070	9.747	± 0.368
REML-LOG	0.440**	± 0.012	0.151**	± 0.007	0.421**	± 0.020	0.157**	± 0.008
Poisson-truncated	2.119	± 0.027	na		2.272**	± 0.025	na	
REML-0	1.657**	± 0.060	na		1.753**	± 0.053	na	

^a See text; ** different ($P < 0.01$) from true value based on *t*-test; na: not applicable.

Empirical mean biases of predictions of breeding values for dams and daughters are shown in table VI for $\sigma_u^2 = 0.056$ and table VII for $\sigma_u^2 = 0.125$. Poisson-based methods and BLUP gave unbiased estimates of breeding values while BLUP-LOG and BLUP-0 performed poorly; BLUP-LOG had a downward bias and BLUP-0 had an upward bias. Predictions of breeding values for the truncated Poisson analysis were generally unbiased.

Empirical MSEs of predictions of breeding values are shown in tables VIII ($\sigma_u^2 = 0.056$) and IX ($\sigma_u^2 = 0.125$). Paired *t*-tests were used in assessing the performance of the comparisons Poisson *versus* BLUP (and BLUP-LOG) and Poisson truncated *versus* BLUP-0. BLUP-LOG and BLUP-0 procedures had the largest MSEs, probably due to their substantial empirical bias. For $\sigma_u^2 = 0.056$ (table VIII), the Poisson procedure and BLUP had a similar MSE. However, Poisson had a slightly smaller ($P < 0.10$) MSE of prediction of dams' breeding values

Table V. Empirical (30 replicates) mean squared errors of 'average' residual variance relative to Poisson analysis (Poisson = 1) in the one-way model.

<i>Analysis</i> ^a	$\sigma_u^2 = 0.056$		$\sigma_u^2 = 0.125$	
	$E(\lambda_i) \approx 2$	$E(\lambda_i) \approx 8$	$E(\lambda_i) \approx 2$	$E(\lambda_i) \approx 8$
Poisson	1.00	1.00	1.00	1.00
REML	8.85	45.1	14.6	100
REML-LOG	223	1 272	288	1 529
Poisson-truncated	1.27	na	1.86	na
REML-0	26.2	na	31.57	na

^a See text ; na: not applicable.

Table VI. Empirical mean biases of predictions of breeding values (\pm standard error of the mean bias) on the observed scale in the one-way model ($\sigma_u^2 = 0.056$).

<i>Analysis</i> ^a	$E(\lambda_i) \approx 2$		$E(\lambda_i) \approx 8$	
	<i>Dams</i>	<i>Daughters</i>	<i>Dams</i>	<i>Daughters</i>
Poisson	-0.023 \pm 0.015	0.016 \pm 0.014	0.016 \pm 0.049	0.012 \pm 0.041
BLUP	-0.025 \pm 0.015	0.018 \pm 0.014	0.027 \pm 0.049	0.007 \pm 0.041
BLUP-LOG	-0.382 \pm 0.011	-0.345** \pm 0.013	-0.668** \pm 0.046	-0.683** \pm 0.038
Poisson-truncated	-0.020 \pm 0.021	0.015 \pm 0.019		
BLUP-0	0.233* \pm 0.017	0.266** \pm 0.015		

^a See text ; ** different ($P < 0.01$) from true value based on *t*-test of 30 replications.

Table VII. Empirical mean biases of predictions of breeding values (\pm standard error of the mean bias) on the observed scale in the one-way model ($\sigma_u^2 = 0.125$).

<i>Analysis</i> ^a	$E(\lambda_i) \approx 2$		$E(\lambda_i) \approx 8$	
	<i>Dams</i>	<i>Daughters</i>	<i>Dams</i>	<i>Daughters</i>
Poisson	-0.010 \pm 0.017	-0.018 \pm 0.018	-0.066 \pm 0.054	0.025 \pm 0.033
BLUP	-0.008 \pm 0.018	-0.020 \pm 0.021	-0.042 \pm 0.058	0.015 \pm 0.032
BLUP-LOG	-0.405** \pm 0.013	-0.435** \pm 0.021	-0.829** \pm 0.060	-0.779* \pm 0.042
Poisson-truncated	0.024 \pm 0.021	0.006 \pm 0.023		
BLUP-0	0.249** \pm 0.019	0.232** \pm 0.021		

^a See text ; ** different ($P < 0.01$) from zero based on a *t*-test of 30 replications.

when $E(\lambda_i) \approx 2$, and a smaller ($P < 0.05$) MSE for predicting daughters' breeding values when $E(\lambda_i) \approx 8$. For $\sigma_u^2 = 0.125$ (table IX), Poisson and BLUP had similar MSEs when $E(\lambda_i) \approx 2$, but Poisson had smaller MSEs than BLUP ($P < 0.05$) for both dams and daughters when $E(\lambda_i) \approx 8$. These small differences between BLUP and Poisson are somewhat surprising in view of the different scales of prediction, and their practical significance is an open question. It was also surprising that the differences between BLUP and Poisson tended to be more significant with $E(\lambda_i) \approx 8$ than with $E(\lambda_i) \approx 2$, since it is known that the Poisson distribution approaches a normal distributions as λ_i increases (Haight, 1967). However, this may be due to a higher power of the test when detecting a larger difference. Another explanation may be that a lower $E(\lambda_i)$ leads to a lower 'pseudoheritability' and, hence, a lower degree of association between phenotypes and breeding values. In this case, the linear and Poisson models may differ less when predicting breeding values because of a higher degree of shrinkage towards zero. When counts of zero were excluded, the Poisson-truncated method had always smaller MSEs of prediction of breeding values than the BLUP-0 method.

Table VIII. Empirical mean squared errors of predictions of breeding values of the observed scale relative to Poisson analysis (Poisson = 1) in the one-way model ($\sigma_u^2 = 0.056$).

<i>Analysis</i> ^a	$E(\lambda_i) \approx 2$		$E(\lambda_i) \approx 8$	
	<i>Dams</i>	<i>Daughters</i>	<i>Dams</i>	<i>Daughters</i>
Poisson	1.00	1.00	1.00	1.00
BLUP	1.00 [†]	1.00	1.00	1.01*
BLUP-LOG	1.68**	1.57**	1.22**	1.22**
Poisson-truncated	1.05	0.98		
BLUP-0	1.35**	1.37**		

^a See text; differences between Poisson *versus* BLUP (or BLUP-LOG) and Poisson-truncated *versus*; BLUP-0 were assessed using paired *t*-tests: [†] $P < 0.10$; * $P < 0.05$; ** $P < 0.01$.

Mixed model

Because variance components estimated by MML and REML are on different scales, empirical coefficients of variation (CV) were used to provide a basis of comparison (fig 1). Clearly, REML estimates were more variable than their Poisson counterparts. No clear pattern with respect to increasing values of the variance components emerged, except that CVs for both MML service sire and genetic estimates seemed relatively more stable while CVs for REML service sire estimates steadily increased for values of σ_u^2 larger than 0.0250 (*ie* $\sigma_s^2 = 0.0500$). For count data with low means, variance components that are estimated under linear mixed-effects models, more general than the one-way random effects model, are virtually meaningless. In Poisson-generated data, these components are highly

Table IX. Empirical mean squared errors of predictions of breeding values of the observed scale relative to Poisson analysis (Poisson = 1) in the one-way model ($\sigma_u^2 = 0.125$).

Analysis ^a	$E(\lambda_i) \approx 2$		$E(\lambda_i) \approx 8$	
	Dams	Daughters	Dams	Daughters
Poisson	1.00	1.00	1.00	1.00
BLUP	1.00	1.00	1.03	1.05*
BLUP-LOG	1.39**	1.44**	1.27**	1.26**
Poisson-truncated	1.04	1.01		
BLUP-0	1.22**	1.15**		

^a See text; differences between Poisson *versus* BLUP (or BLUP-LOG) and Poisson-truncated *versus* BLUP-0 were assessed using paired *t*-tests: * $P < 0.05$; ** $P < 0.01$.

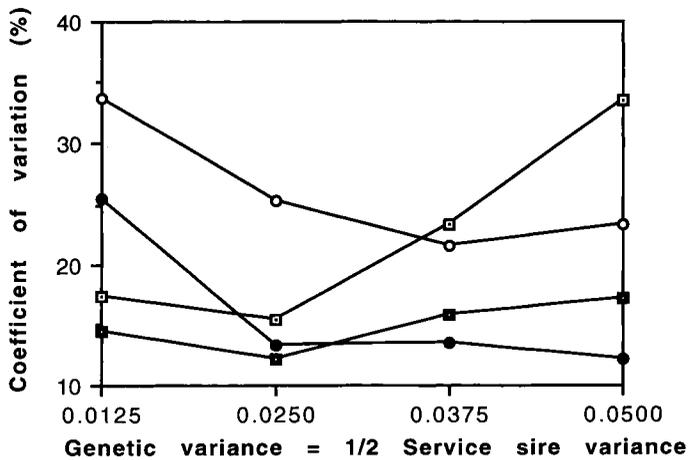


Fig 1. Coefficients of variation for variance component estimates in the mixed-model simulation. —□— : REML SS; —○— : REML GENETIC; —■— : MML SS; —●— : MML GENETIC; SS = service sire variance component; GENETIC = genetic variance component.

heteroskedastic from one observation to the next, depending on both experimental design and location parameters (Foulley and Im, 1993).

Relative biases of the MML estimates are given in figure 2. Relative biases were less than 4% for all 4 sets of variance components, with no clear trend with respect to the true values of variance components. Using *t*-tests, these biases did not differ from zero. Unlike results obtained with threshold models (*eg*, Höschele *et al*, 1987; Simianer and Schaeffer, 1989), a small subclass (equal to 1 in the Poisson animal model) did not lead to detectable bias of variance component estimates in the Poisson mixed model.

Relative errors (square roots of the empirical MSEs, expressed as a percentage of the true variance component values) are given in figure 3. Trends with respect to

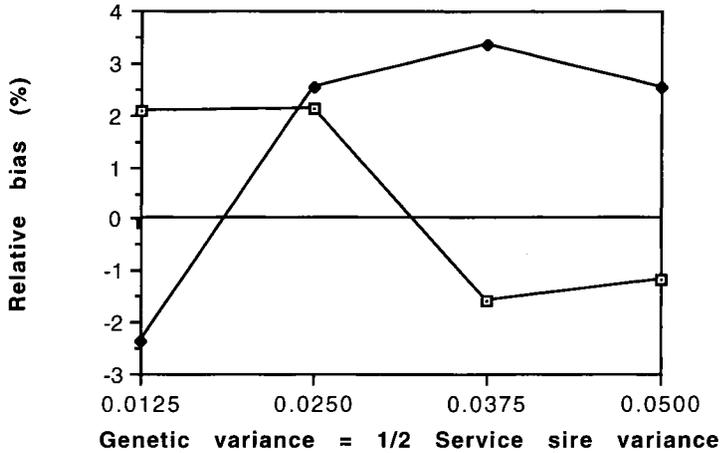


Fig 2. Relative bias (%) of MML variance component estimates in the mixed-model simulation. —□— : Animal; —●— : Sire.

the size of the true variances were somewhat opposite for service sire and genetic variance component estimates. Relative error for the genetic component decreased as the true variance increased, whereas the error of the service sire estimates increased somewhat with the true value of the parameter. The relative errors of MML estimates were almost identical to their empirical CVs (fig 1) because of the small bias, as shown in figure 2.

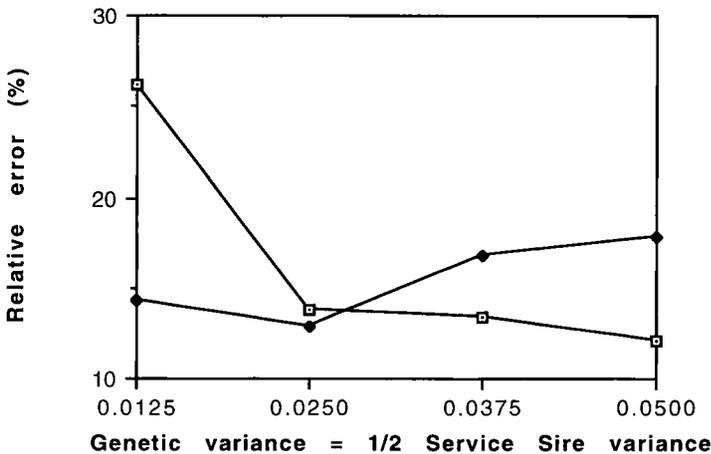


Fig 3. Relative error of MML variance component estimates in the mixed-model simulation. —□— : Animal; —●— : Sire.

Empirical biases of fixed-effect estimates obtained with linear and Poisson procedures are given in figures 4-7 for the 4 different sets of true variance

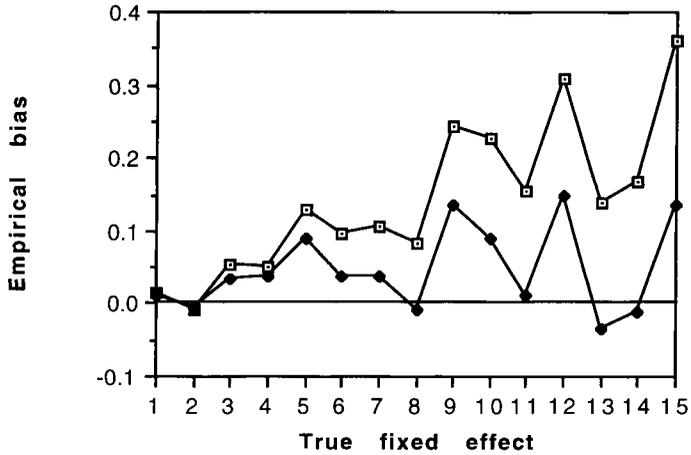


Fig 4. Empirical bias of BLUE and Poisson estimates of fixed effects in the mixed-model simulation: variance component set A ($\sigma_u^2 = 0.0125; \sigma_s^2 = 0.0250$). —□— : BLUE; —●— : Poisson.

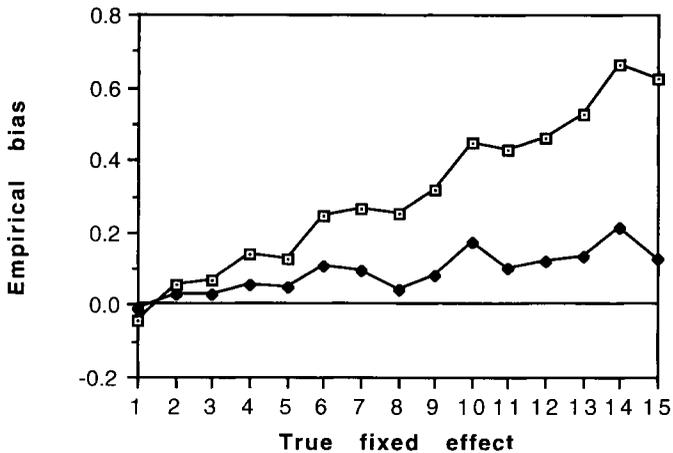


Fig 5. Empirical bias of BLUE and Poisson estimates of fixed effects in the mixed-model simulation: variance component set B ($\sigma_u^2 = 0.0250; \sigma_s^2 = 0.0500$). —□— : BLUE; —●— : Poisson.

components. The 2 methods gave estimates that were biased upwards, but the bias was larger for BLUE in all 4 cases. This apparent paradox can be explained by the fact that BLUE is an unbiased estimator of:

$$E(\lambda_{ijk}) = \exp(\mu + H_i) \exp\left(\frac{\sigma_u^2 + \sigma_s^2}{2}\right)$$

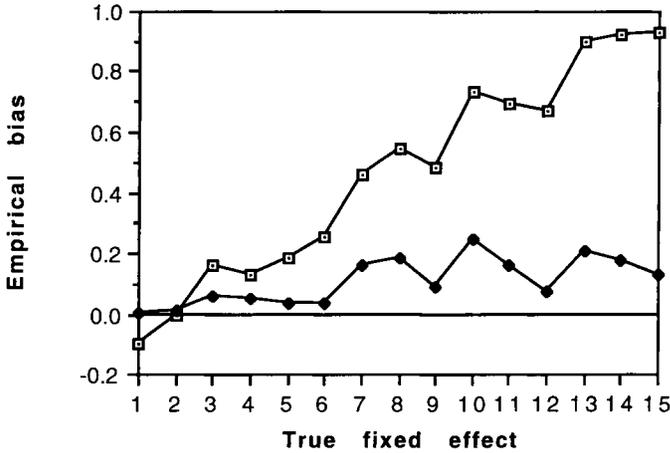


Fig 6. Empirical bias of BLUE and Poisson estimates of fixed effects in the mixed-model simulation: variance component set C ($\sigma_u^2 = 0.375$; $\sigma_s^2 = 0.0750$). —□— : BLUE; —●— : Poisson.

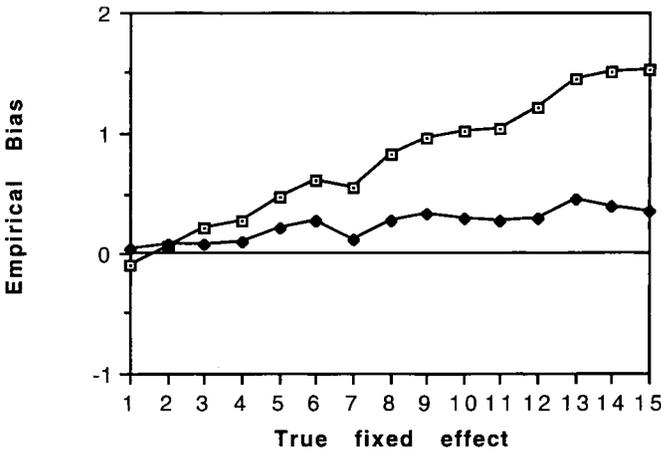


Fig 7. Empirical bias of BLUE and Poisson estimates of fixed effects in the mixed-model simulation: variance component set D ($\sigma_u^2 = 0.0500$; $\sigma_s^2 = 0.1000$). —□— : BLUE; —●— : Poisson.

and not $\exp(\mu + H_i)$. Empirical biases for the 2 methods, and their difference, increased with increasing values of fixed effects and with higher values of variance components. The upward biases of the Poisson estimates were generally stable across sets of variance components, being always less than 0.5. However, the bias of the BLUEs of the fixed effects increased substantially as variance increased. Although Poisson estimates of fixed effects were manifestly biased upwards at higher embryo yields, the magnitude of their bias was several times smaller than that of

BLUE estimates. Empirical MSEs of the fixed effects estimates are depicted in figures 8–11 for each of the 4 sets of variance components. As with empirical biases, MSEs of the linear model and Poisson estimates, and the differences between the MSEs of the 2 procedures tended to increase with increasing values of the variance components and with the true values of the fixed effects. The MSEs of the Poisson estimates were again more stable across sets of variance components and, although tending to increase with the value of the fixed effects, were much smaller than those of BLUE estimates. For example (fig 11), when embryo yield was around 15, the MSE of BLUE was about 6 times larger than that of Poisson estimates.

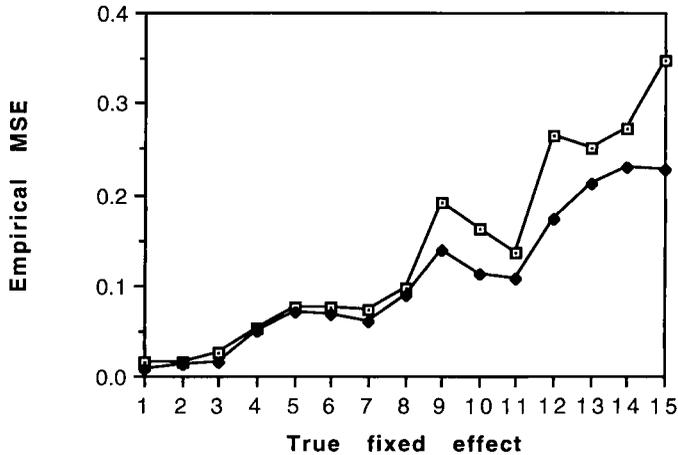


Fig 8. Empirical mean squared errors of BLUE and Poisson estimates of fixed effect solutions in the mixed-model simulation: variance component set A ($\sigma_u^2 = 0.0125$; $\sigma_s^2 = 0.0250$). —□— : BLUE; —●— : Poisson.

Because the values of fixed effects were constructed such that:

$$\exp(\mu + H_{i+1}) - \exp(\mu + H_i) = 1; \quad i = 1, 2, \dots, 14$$

it was of interest to examine the extent to which the 2 estimators would capture such difference. If the estimates are regressed against true values, the ‘best’ estimator should give a slope close to one. Such regressions were assessed by ordinary least-squares. Empirical biases and MSEs of the estimates of the regression effects are given in figures 12 and 13, respectively. Regression estimates obtained under both models were slightly biased downwards, particularly BLUE; the absolute bias increased somewhat as true variance components increased in value. The differences between the biases of the 2 estimators were found significant in all cases using a paired *t*-test ($P < 0.05$ for variance component set A, and $P < 0.01$ for variance component sets B, C, D). The differences in empirical MSEs were in the same direction, *ie* BLUE had larger MSEs.

Empirical average biases and empirical average MSEs of the predictions of service sire effects are given in figures 14 and 15, respectively. The statistics were slightly in favor of the Poisson model. Both methods had empirical average biases different

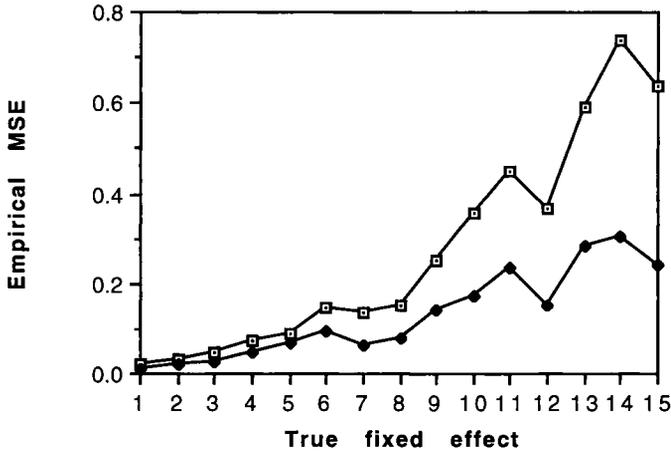


Fig 9. Empirical mean squared errors of BLUE and Poisson estimates of fixed effect solutions in the mixed-model simulation: variance component set B ($\sigma_u^2 = 0.0250$; $\sigma_s^2 = 0.0500$). —□— : BLUE; —◆— : Poisson.

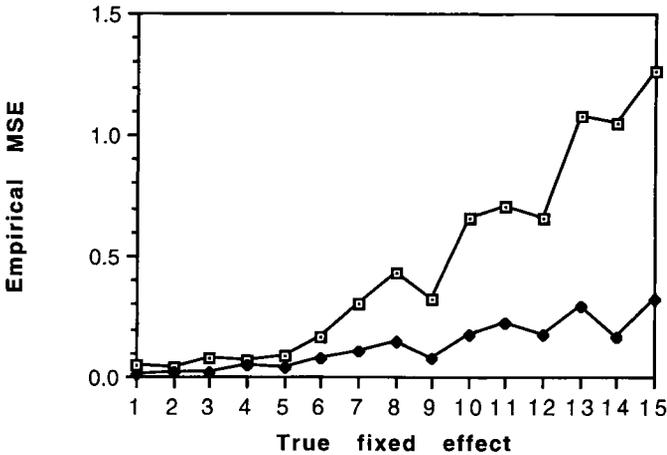


Fig 10. Empirical mean squared errors of BLUE and Poisson estimates of fixed effect solutions in the mixed-model simulation: variance component set C ($\sigma_u^2 = 0.0375$; $\sigma_s^2 = 0.0750$). —□— : BLUE; —◆— : Poisson.

($P < 0.05$) from 0 only at $\sigma_s^2 = 0.0500$. Based on paired t -tests, empirical average MSEs were not different between the 2 methods.

Statistics associated with prediction of genetic effects are depicted in figures 16 and 17, with results presented separately for base generation sires and dams, and for their female progeny surviving to age at first breeding. Poisson-predicted breeding values tended to be biased downwards, whereas BLUP predictions were biased upwards; however, average bias was smaller for Poisson predictions. BLUP sire

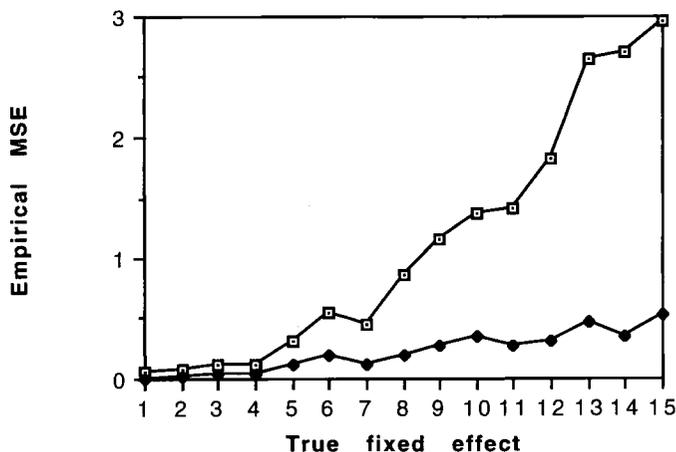


Fig 11. Empirical mean squared errors of BLUE and Poisson estimates of fixed effect solutions in the mixed-model simulation: variance component set $D(\sigma_u^2 = 0.0500; \sigma_s^2 = 0.1000)$. —□— : BLUE; —●— : Poisson.

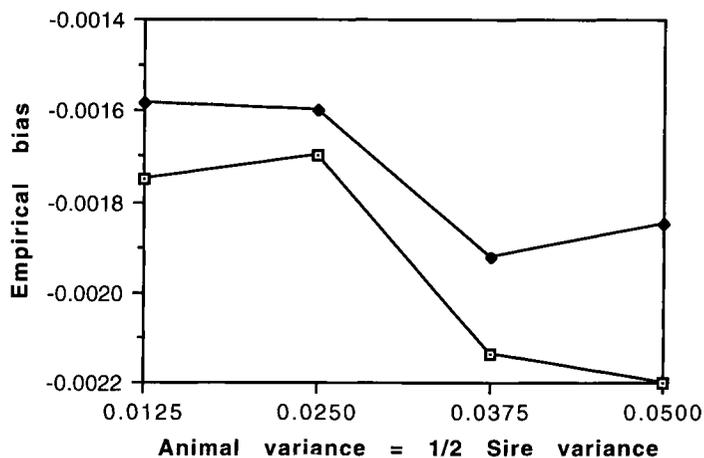


Fig 12. Empirical bias of estimates of the 'regression' of BLUE and Poisson solutions on true values in the mixed-model simulation. —□— : BLUE; —●— : Poisson.

solutions were significantly biased ($P < 0.01$) at the 2 highest variance components. Dam and daughter BLUP solutions were significantly biased ($P < 0.01$ in all cases, except $P < 0.05$ for dams at $\sigma_u^2 = 0.0125$). Poisson dam solutions were biased ($P < 0.01$) only at $\sigma_u^2 = 0.0250$, while Poisson daughter solutions were biased at $\sigma_u^2 = 0.0250$ ($P < 0.01$) and $\sigma_u^2 = 0.0375$ ($P < 0.05$). Certainly, all predictions reflect uncertainty in the baseline estimates of fixed effects given in [17] and [20] for Poisson and BLUP models, respectively, and upward biases in BLUP predictions of random effects reflect upward biases for the baseline estimates (fig 4-7).

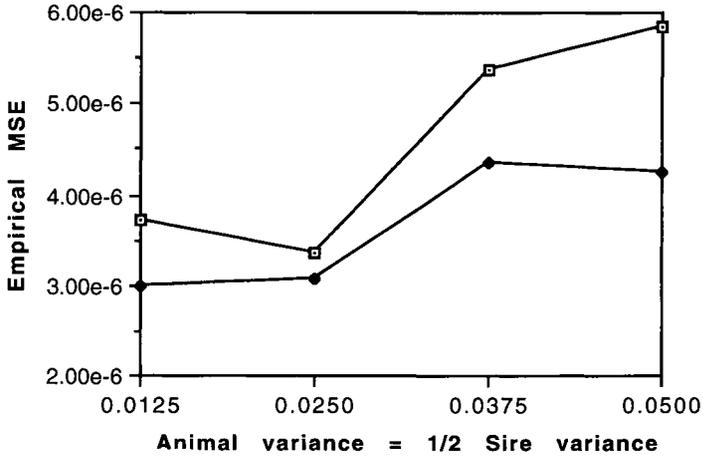


Fig 13. Empirical mean squared errors of estimates of the 'regression' of BLUE and Poisson solutions on true values in the mixed-model simulation. —□— : BLUE; —●— : Poisson.

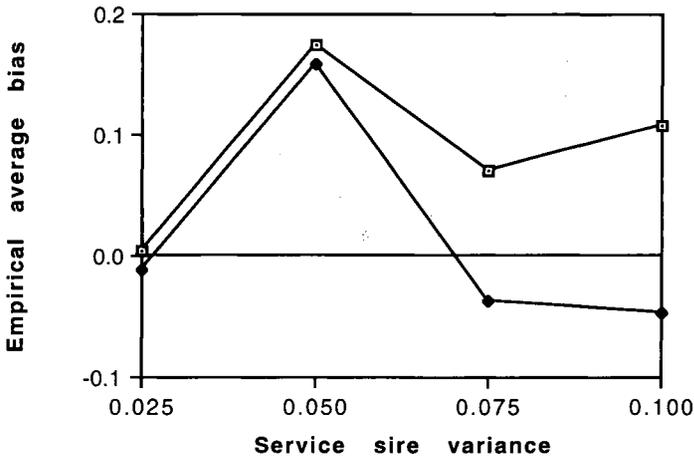


Fig 14. Empirical average bias of prediction of service sire effects in the mixel-model simulation. —□— : BLUP; —●— : Poisson.

Differences in empirical average MSEs of prediction of breeding values between the 2 procedures are depicted in figure 17. Although the differences between the 2 methods were small, a paired *t*-test gave significant differences for daughters ($P < 0.01$ for $\sigma_u^2 = 0.0125, 0.0375,$ and 0.0500 ; $P < 0.05$ for $\sigma_u^2 = 0.0250$) and for base generation dams, except at $\sigma_u^2 = 0.0125$ ($P < 0.01$ for $\sigma_u^2 = 0.0250, 0, 0500$; $P < 0.05$ for $\sigma_u^2 = 0.0375$). Differences in empirical average MSE of predictions of base generation sires were not significant.

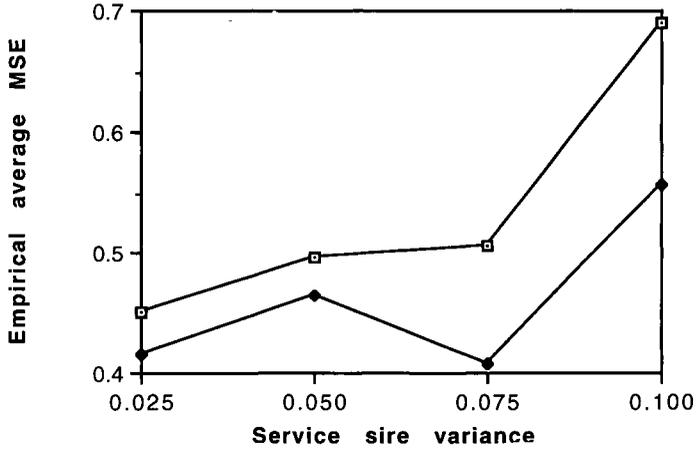


Fig 15. Empirical average mean squared errors of prediction of service sire effects in the mixed-model simulation. —□— : BLUP; —●— : Poisson.

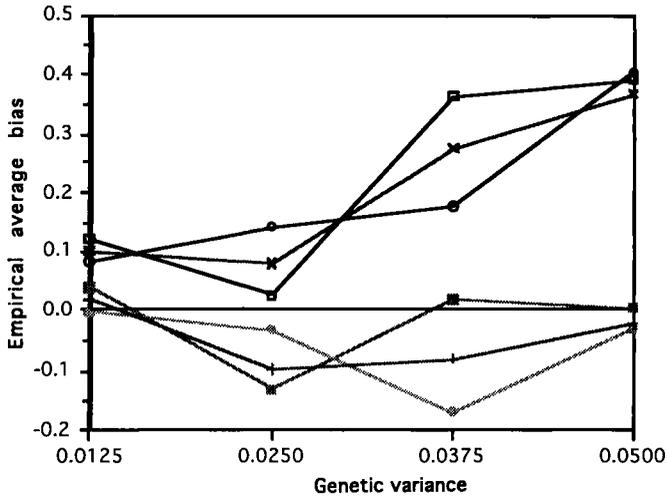


Fig 16. Empirical average bias of prediction of genetic effects in the mixed-model simulation. —□— : BLUP sire; —○— : BLUP dam; —×— : BLUP daughter; —■— : Poisson sire; —●— : Poisson dam; —▲— : Poisson daughter.

CONCLUSIONS

This study evaluated the sampling performance of estimators of location and dispersion parameters in Poisson mixed models, and of predictors of breeding values in the context of simulated MOET schemes. Records on embryo yield were drawn from Poisson distributions for 4 populations characterized by appropriate parameter values. Analyses were carried out with the Poisson model and with a linear model in each of these populations.

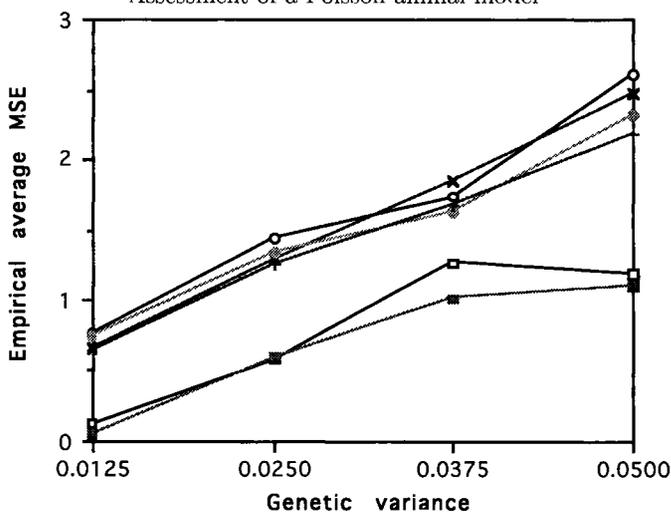


Fig 17. Empirical average mean squared error of prediction of genetic effects in the mixed-model simulation. —□— : BLUP sire; —○— : BLUP dam; —×— : BLUP daughter; —■— : Poisson sire; —●— : Poisson dam; —+— : Poisson daughter.

In general, the estimates of parameters and the predictions of random effects obtained with the Poisson model were better than their linear model counterparts. This is not surprising because, to begin with, in the Poisson analysis, the data were analyzed with the models employed to generate the records in the simulation. Further, if the distribution is indeed Poisson, there is no appropriate variance decomposition independent of location parameters in the linear model, which makes the variance component estimators employed with normal data somewhat inadequate for estimating dispersion parameters, particularly in the mixed effects model. A log-transformation improved (relative to untransformed REML) the mean squared error performance of REML estimates of genetic variance, but worsened the estimates of residual variance. Similarly, fixed effects were estimated by BLUE with a larger bias and mean squared error than when estimated by the Poisson model.

BLUP was found to be robust in predicting random effects, although Poisson joint modes were significantly better with respect to bias and MSE in many cases. This is remarkable, considering that the BLUP predictors were based on location-parameter-dependent estimates of variance. BLUE, however, showed bias with increasing values of fixed effects and variance components. Truncated-Poisson estimators and predictors always outperformed BLUP and REML when zero counts were excluded from the analysis; truncation of null counts may be common in field records on traits such as litter size (Perez-Enciso *et al.*, 1993).

Estimates of variance components obtained by MML in Poisson mixed models did not exhibit the typical bias due to small subclass sizes encountered often in threshold models. Rather, the downward biases of MML found in the one-way models may be due to small amount of statistical information. After all, MML is, like REML, a biased estimator. However, it should be consistent, because all Bayesian estimators are so under certain forms of selection (Fernando and Gianola, 1986). Höschele

et al (1987) attributed the 'small subclass' bias in threshold models to inadequacy of a normal approximation invoked in MML. The same approximation is made when estimating variance with the Poisson model, but its consequences may be less critical here.

In conclusion, REML and BLUE do not perform well when the assumption of a Poisson distribution holds, even if all pertinent factors in the model are included in the analysis. BLUP, however, was found to be robust, although it had a slightly inferior sampling performance. This study did not address the situation when the data are counts, but the distribution is not Poisson. For this instance, estimators based on the assumption of normality may perform better than those that rely on the Poisson distribution, due to central limit theory. It should also be noted that if the operational Poisson model fails to consider all pertinent explanatory factors, the conditional variance may be substantially greater than the conditional mean of an observation; these 2 parameters are defined to be equal as in [2]. Possible diagnostics for investigating this source of overdispersion were discussed by Dean (1989).

ACKNOWLEDGMENTS

The authors thank M Perez-Enciso and I Misztal for making available to them sparse matrix subroutines for computations involving relationship matrices. The authors also thank R Fernando and T Wang for helpful suggestions and comments. The work was partially funded by grants from 21st Century Genetics (Shawano, WI, USA), Sire Power Inc (Tunkhannock, PA, USA) and by the National Pork Producers Council (Des Moines, IO, USA).

REFERENCES

- Boldman KG, Van Vleck LD (1991) Derivative-free restricted maximum likelihood estimation in animal models with a sparse matrix solver. *J Dairy Sci* 74, 4337-4343
- Carriquiry AL, Gianola D, Fernando RL (1987) Mixed model analysis of a censored normal distribution with reference to animal breeding *Biometrics* 43, 929-939
- Dean CB (1992) Testing for overdispersion in Poisson and binomial regression models. *J Am Stat Assoc* 87, 451-457
- Eisenstat SC, Gursky MC, Schultz MH, Sherman AH (1982) Yale sparse matrix package. I. The symmetric codes. *Int J Numer Meth Eng* 18, 1145-1151
- Fernando RL, Gianola D (1986) Optimal properties of the conditional mean as a selection criterion. *Theor Appl Genet* 72, 822-825
- Foulley JL, Gianola D, Im S (1987) Genetic evaluation of traits distributed as Poisson-binomial with reference to reproductive characters. *Theor Appl Genet* 73, 870-877
- Foulley JL, Gianola D, Im S (1990) Genetic evaluation for discrete polygenic traits in animal breeding. In: *Advances in Statistical Methods for Genetic Improvement of Livestock* (D Gianola, K Hammond, eds) Springer Verlag, Heidelberg, 361-409
- Foulley JL, Im S (1993) A marginal quasi-likelihood approach to the analysis of Poisson variables with generalized linear mixed models. *Genet Sel Evol* 25, 101-107

- Freeman AE (1986) Genetic control of reproduction and lactation of dairy cattle. *In: Proc 3rd World Congress on Genetics Applied to Livestock Production*, Lincoln, NE 16-22 July 1986 (GE Dickerson, RK Johnson, eds) vol XI, 3-13
- Gianola D, Foulley JL (1983) Sire evaluation for ordered categorical data with a threshold model. *Genet Sel Evol* 15, 201-224
- Graser HU, Smith SP, Tier B (1987) A derivative-free approach for estimating variance components in animal models by restricted maximum likelihood. *J Anim Sci* 64, 1362-1370
- Hahn J (1992) Attempts to explain and reduce variability of superovulation. *Theriogenology* 38, 269-275
- Haight FA (1967) *Handbook of the Poisson Distribution* Wiley, New York
- Harville DA, Mee RW (1984) A mixed model procedure for analyzing ordered categorical data. *Biometrics* 40, 398-408
- Hasler JF (1992) Current status and potential of embryo transfer and reproductive technology in dairy cattle. *J Dairy Sci* 75, 2857-2779
- Höschle I, Gianola D, Foulley JL (1987) Estimation of variance components with quasi-continuous data using Bayesian methods. *J Anim Breed Genet* 104, 334-349
- Keller DS, Teepker G (1990) The effect of variability in responses to superovulation on donor cow selection differentials in nucleus breeding schemes. *J Dairy Sci* 73, 549-554
- Kincaid DR, Respass JR, Young DM, Grimes RG (1982) Algorithm 586 ITPACK 2C: A Fortran package for solving sparse linear systems by adaptive accelerated methods *ACM Trans Meth Software* 8, 302-322
- Liboriussen T, Christensen LG (1990) Experiences from implementation of a MOET breeding scheme for dairy cattle. *In: Proc 4th World Congress on Genetics Applied to Livestock Production* (WG Hill, R Thompson, JA Woolliams, eds) vol XIV, 66-69
- Lohuis M, Smith C, Burnside EB (1990) Embryo transfer results in a dispersed nucleus MOET/AI testing scheme. *In: Proc 4th World Congress on Genetics Applied to Livestock Production* (WG Hill, R Thompson, JA Woolliams, eds) vol XIV, 221-224
- McCullagh P, Nelder JA (1989) *Generalized Linear Models: 2nd edn*, Chapman and Hall, London
- Meijering A, Gianola D (1985) Linear versus nonlinear methods of sire evaluation for categorical traits: a simulation study. *Genet Sel Evol* 17, 115-132
- Meuwissen THE (1991) The use of increased female reproductive rates in dairy cattle breeding schemes. *Anim Prod* 52, 21-31
- Morris CA, Day AM (1986) Potential for genetic twinning in cattle. *In: Proc 3rd World Congress on Genetics Applied to Livestock Production*, Lincoln, NE 16-22 July 1986 (GE Dickerson, RK Johnson, eds) vol XI, 14-29
- Nicholas FW, Smith C (1983) Increased rates of genetic change in dairy cattle by embryo transfer and splitting. *Anim Prod* 36, 341-353
- Quaas RL, Zhao Y, Pollak EJ (1988) Describing interactions in dystocia scores with a threshold model. *J Anim Sci* 66, 396-399
- Perez-Enciso M, Tempelman RJ, Gianola D (1993) A comparison between linear and Poisson mixed models for litter size in Iberian pigs. *Livest Prod Sci* 35, 303-316
- Ron M, Ezra E, Weller JI (1990) Genetic analysis of twinning rate in Israeli Holstein cattle. *Genet Sel Evol* 22, 349-359

- Ruane J (1991) The importance of family sizes in adult MOET nucleus breeding schemes in dairy cattle. *Anim Prod* 52, 33-47
- Simianer H, Schaeffer LR (1989) Estimation of variance components between one continuous and one binary trait. *Genet Sel Evol* 21, 303-315
- Tempelman RJ, Gianola D (1993) Marginal maximum likelihood estimation of variance components in Poisson mixed models using Laplacian integration. *Genet Sel Evol* 25, 305-319
- Van Raden PM, Lawlor TJ, Short TH, Hoeschele I (1992) Use of reproductive technology to estimate variances and predict effects of gene interactions *J Dairy Sci* 75, 2892-2901