

Kansas State University Libraries

New Prairie Press

Conference on Applied Statistics in Agriculture

1998 - 10th Annual Conference Proceedings

RISK FACTORS ASSOCIATED WITH CULLING AGE IN DAIRY CATTLE: APPLICATIONS OF FRAILTY MODELS

Mohamed M. Shoukri

Jan M. Sargeant

Follow this and additional works at: <https://newprairiepress.org/agstatconference>



Part of the [Agriculture Commons](#), and the [Applied Statistics Commons](#)



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License](#).

Recommended Citation

Shoukri, Mohamed M. and Sargeant, Jan M. (1998). "RISK FACTORS ASSOCIATED WITH CULLING AGE IN DAIRY CATTLE: APPLICATIONS OF FRAILTY MODELS," *Conference on Applied Statistics in Agriculture*. <https://doi.org/10.4148/2475-7772.1277>

This is brought to you for free and open access by the Conferences at New Prairie Press. It has been accepted for inclusion in Conference on Applied Statistics in Agriculture by an authorized administrator of New Prairie Press. For more information, please contact cads@k-state.edu.

**RISK FACTORS ASSOCIATED WITH CULLING AGE IN DAIRY CATTLE:
APPLICATIONS OF FRAILTY MODELS**Mohamed M. Shoukri ¹ and Jan M. Sargeant ²¹ Department of Population Medicine, University of Guelph,
Guelph, Ontario, Canada N1G 2W1² Food Animal Health and Management Center, College of Veterinary Medicine,
Kansas State University, Manhattan, Kansas 66506**ABSTRACT**

Culling decisions for dairy cattle are an important component of dairy herd management. To investigate risk factors for culling, farms (clusters) constitute the sampling units. Therefore, we believe that ages-at-culling may be correlated within farms. The score test on the null hypothesis of no extra-variation in survival data was not supported by age-at-culling data collected from 72 dairy farms from the province of Ontario, Canada. To correct for the intraherd correlation, three modelling approaches were used to fit the data: Population-Averaged (PA), cluster-specific (CS), and Random Effects Models (RAEM). The modelling approaches are described and compared using the dairy cow culling data.

1. INTRODUCTION

“Culling” is defined as the removal of animals from a herd. For dairy production enterprises, the decision to cull provides the manager with the opportunity to achieve progress in genetic selection and to select animals for culling which are diseased or reproducing suboptimally. However, culled cows can represent a sizable loss to the dairy herd. The cost associated with replacement heifers averages 20% of the overall operating expenses of a dairy herd (Fetrow, 1988). The decision to cull is controlled by many factors, such as the long term economic strategy of the dairy enterprise, udder health and disease control strategies, reproductive

management, and planning for genetic improvement.

Culling may be described as a binary outcome (culled / not culled) or by using age-at-culling as a response variable. The advantage of age-at-culling as a response variable is that it incorporates more information. Therefore, instead of modelling a binary outcome where logistic regression is traditionally used, we deal with models for the analysis of failure time data. By failure, we mean the occurrence of a pre-specified event. By failure time, we mean the period of time taken for the event to occur. In this study, the event is the removal of a cow from the dairy herd, and the failure time is the age-at-culling.

An important feature of failure time data is the possibility that the data may be censored. This refers to the circumstance where some animals are event free (have not been culled at the termination of the study) or were removed for reasons other than culling.

Most of the statistical models and methods for failure time data were developed under the assumption that the observations from subjects are statistically independent of each other. While the assumption of independence may be valid in many applications, it has become evident that this assumption is violated in other fields of applications. In our study, and since our sampling unit is a farm (herd), there is every reason to expect strong within herd correlation among culling ages.

There is little experience with the practical problems which arise in the use of correlated event time models and their inferential procedures in the field of agriculture. Therefore, the objective of this paper was to describe and compare several available modelling techniques.

2. THE DATA

The data were from 72 Ontario farms involved in a 2-year observational study (Sargeant,

1996). The data used in this analysis were collected between April 1993 and October, 1994. Production and culling data were obtained from the Ontario Dairy Herd Improvement corporation. Disease data were recorded by the producers throughout the study. Culling age was calculated as the date of birth to the date the cow was culled. The definition of culling was restricted to the removal of cows by slaughter. During the study, cows that were in the herd and removed for other reasons (e.g. sold to other dairy farms) were considered to be lost-to-follow up and their failure time was treated as a censored observation. The analyses used three explanatory variables which were believed to be prognostic indicators for culling: parity, cumulative milk production in the previous lactation, and the presence or absence of mastitis in the lactation period prior to the lactation in which the cow was culled.

Research has shown that low-producing cows are more likely to be culled, and that dairy cows have an additional 2.3 months of total life herd and 1.5 additional lactations in the herd for each 1000 l (450 kg) of first lactation milk production above the herd average (Greer, et al. 1980). Parity, which represents the number of times a cow has calved, was perceived to be an important covariate. Dairy cattle are bred to calf for the first time at approximately two years of age. Following the initial calving, the time period between each subsequent calving consists of a calving-to-conception period (which varies between farms and cows within farms), and a 283 day gestation period which is fairly constant among cows. Therefore, as the number of parity periods increases, the chronological age of the cow will also increase. The third covariate, considered as a potential risk factor was the presence or absence of mastitis (inflammation of the udder).

At the end of the study period there were 1193 failures and 130 censored observations (a

random sub-sample of all censored observations chosen to approximately equal 10% of the number of failures). Failure from causes other than culling was treated as right censoring and therefore assumed to be noninformative. This assumption allowed us to focus upon the methodological issues involved in modelling correlated failure time data.

3. TESTING FOR HETEROGENEITY

Let T_{ij} denote the failure time of the j th cow within the i th herd (cluster) ($j=1,2,\dots,n_i$; $i=1,2,\dots,k$) where n_i is the herd size, and k is the number of herds in the sample. Moreover, let δ_{ij} be a censoring indicator taking the value 1 if T_{ij} is not censored and 0 otherwise. Thus for each individual in each cluster we observe the triple $\{(T_{ij}, \delta_{ij}, X_{ij}): 1 \leq j \leq n_i; i=1,2,\dots,k\}$ with X_{ij} being the covariate vector. A common approach for modelling correlated survival data is to assume that the within cluster association is induced by a shared frailty, a term that represents a set of unobserved or latent variables that cause extra variation in the response variable. A multivariate model for the observed data is then obtained by averaging over an assumed distribution for the latent variable. Let ξ_i denote the unobserved frailty for the i th cluster. We assume a Cox-proportional hazard regression (PH). Following Clayton and Cuzick (1985), we assume that ξ_i act multiplicatively on the Cox PH. Since T_{ij} are conditionally independent (conditional on the frailty), the contribution of the i th cluster to the likelihood function is given by

$$L_i(\beta | \xi_i) = \prod_{i=j}^{n_i} (\lambda(t_{ij} | \xi_i))^{\delta_{ij}} S(t_{ij} | \xi_i) \tag{3.1}$$

where

$$S(t|\xi_i) = \exp\left[-\xi_i \Lambda_o(t) e^{x'_{ij}\beta}\right],$$

$\Lambda_o(t)$ is the base-line cumulative hazard and t denotes the transpose of the covariate matrix.

The conditional likelihood of the i th cluster can be written as

$$L_i(\beta|\xi_i) = w_i e^{-\xi_i \mu_i} (\xi_i \mu_i)^{\delta_i} \tag{3.2}$$

where

$$w_i = \prod_{j=1}^{n_i} \left[\frac{\xi_i \lambda_o(t_{ij}) e^{x'_{ij}\beta}}{\sum_{j=1}^{n_i} \xi_i \Lambda_o(t_{ij}) e^{x'_{ij}\beta}} \right]^{\delta_{ij}},$$

$$\mu_i = \sum_{j=1}^{n_i} \Lambda_o(t_{ij}) e^{x'_{ij}\beta} = \sum_{j=1}^{n_i} \mu_{ij},$$

and

$$\delta_i = \sum_{j=1}^{n_i} \delta_{ij}$$

Specification of the multivariate distribution of the failure times is completed by assuming that the ξ_i are independent variables drawn from a family of distributions, $g(\xi, \theta)$ indexed by an unknown parameter θ . Hence the overall likelihood function is

$$L = \prod_{i=1}^k \left[\int L_i(\beta | \xi_i) g(\xi_i, \theta) d \xi_i \right] = \prod_{i=1}^k \left[E (L_i(\beta | \xi_i)) \right] \quad 3.3$$

(it is assumed that $E(\xi_i)=1$ and $\text{Var}(\xi_i)=\theta < \infty$).

The proposed test for heterogeneity is a score test for the hypothesis $\theta=0$, and is calculated as

$$Z = \frac{S}{\sqrt{\text{Var}(S)}} \quad 3.4$$

where

$$S = \frac{1}{2} \left[\sum_{i=1}^k \left[\sum_{j=1}^{n_i} (\delta_{ij} - \Lambda_o(t_{ij}) e^{x_{ij} \beta}) \right]^2 - \delta_i \right]$$

and

$$\text{Var}(S) = I_{\theta\theta} - I_{\theta\beta} I_{\beta\beta}^{-1} I_{\theta\beta}^T \quad 3.5$$

The elements on the right handside of equation 3.5 are the negatives of the second partial derivatives of the log-likelihood with respect to the indicated parameters, evaluated at their maximum likelihood estimates.

When we reject the null hypothesis $H_0:\theta=0$, the standard errors of the estimated regression parameters obtained from the conventional cox model are likely to be under estimated,

resulting in inflated type I error rates.

4. POPULATION AVERAGE APPROACH

Two marginal approaches for the analysis of correlated failure time observations will be discussed. The first, the Group Jackknife Estimator (GJE), was developed by Lee et al. (1992) and the second is the generalized estimating equation (GEE) of Liang and Zeger (1986).

Group Jackknife Estimator (GJE)

In this approach, a proportional hazards model is used, with a variance adjustment to account for the lack of independence within herd. The GJE of the variance was derived by Lin and Wei (1989) based on Huber's (1967) sandwich estimator $V(\hat{\beta})$ of $\hat{\beta}$:

$$V(\hat{\beta}) = A^{-1} (B^T B) A^{-1} = H^T H \quad 4.1$$

where $H = B A^{-1}$ is applicable to the estimates obtained from the partial likelihood and

$$B_r = \frac{\partial \log PL(\beta)}{\partial \beta_r} = \sum_{i=1}^N \Delta_i \{ x_{ri} - a_{ri} \} \quad r=1,2,\dots,p \quad 4.2$$

where

$$a_{ri} = \frac{\sum_{l \in R_i} x_{rl} \exp(x_l^T \hat{\beta})}{\sum_{l \in R_i} \exp(x_l^T \hat{\beta})}$$

and p is the number of covariates.

$$\begin{aligned}
 A_{rs} &= -\frac{\partial^2 \log PL(\beta)}{\partial \beta_r \partial \beta_s} \\
 &= \sum_{i=1}^N \Delta_i \left[\frac{\sum_{\ell \in R_i} x_{r\ell} x_{s\ell} \exp(x_{\ell}^t \hat{\beta})}{\sum_{\ell \in R_i} \exp(x_{\ell}^t \hat{\beta})} - a_{ri} a_{si} \right] \quad 4.3
 \end{aligned}$$

$r, s = 1, 2, \dots, p$

The Generalized Estimating Equation (GEE):

The GEE likelihood indicates that, conditional on (t_{ij}, β) , the censoring indicator δ_{ij} follows a Poisson distribution with mean $(\xi_i \mu_{ij})$,

$$\mu_{ij} = \Lambda_o(t_{ij}) \exp(x_{ij}^t \beta) \quad 4.4$$

This Poisson likelihood formulation for censored survival data has been exploited by Segal and Neuhaus (1993) and was linked to the “Generalized Estimating” (GEE) approach of Liang and Zeger (1986) for clustered data. Briefly, the data comprise independent vector observations $\delta_i (i=1, 2, \dots, k)$ where δ_i is $(n_i \times 1)$ with mean vector $\mu_i = E(\delta_i)$ and $n_i \times n_i$ covariance matrix $\Sigma_i = \text{Cov}(\delta_i)$. There are two main components required to model data within the GLM framework:

(a) specifying a link function $h_j(\mu_{ij}) = x_{ij}^t \beta$ and $\Sigma_i = \sigma^2 V(\mu_i)$ where x_{ij} are covariate vectors, β is the regression coefficient, $V(\bullet)$ is a matrix of covariance functions and σ^2 is a scalar factor

(b) estimating β by equating to zero the quasi-score function

$$Q = \sum D_i^T V^{-1}(\mu_i) (\delta_i - \mu_i) \quad 4.5$$

where $D_i = \partial \mu_{ij} / \partial \beta_r$ $r=1,2,\dots,q$. The solution of $Q=0$ is generally consistent and asymptotically $N(\beta, V_\beta)$, with $V_\beta = \{\sum_i D_i^T V^{-1}(\mu_i) D_i\}^{-1} \sigma^2$.

In situations where $V(\mu_i)$ is non-diagonal there are parameters additional to (β, σ^2) that need to be estimated. Liang and Zeger (1986) proposed the use of a “working” correlation matrix parameterized in terms of a vector α with an additional estimating equation for α if necessary. They suggested writing Σ_i as $\sigma^2 A_i R_i A_i$ where R_i is the correlation matrix of δ_i and A_i is a diagonal matrix with jj^{th} entry $\sigma \sqrt{V(\delta_{jj})}$.

Segal and Neuhaus (1993) proposed using the GEE Poisson regression with the censoring indicator as the dependent variable, instead of the survival time (which is incorporated as an offset variable). This allows a robust variance estimate of the $\hat{\beta}$ and a moment estimate of α to be readily obtained, an approach which seems quite suitable to analyse clustered survival data with covariates.

The robust standard errors are the square roots of the diagonal elements of the “sandwich estimator” of the covariance matrix of $\hat{\beta}$ which was given by Liang and Zeger (1986) as,

$$ACOV(\hat{\beta}) = U^{-1} \left\{ \sum_{i=1}^k D_i^T \hat{V}_i^{-1} Cov(\delta_i) \hat{V}_i^{-1} D_i \right\} U^{-1} \quad 4.6$$

where

$$U = \sum_{i=1}^k D_i^T \hat{V}_i^{-1} D_i$$

5. CLUSTER-SPECIFIC APPROACH

I- Hierarchical maximum likelihood:

In the conditional proportional hazard model, the interpretation of β is specific to the individual farm. Therefore, β has a conditional interpretation (conditional on the cluster). The same relative hazard for cows from two different farms is $\exp(\beta_k + u_i - u_j)$, ($u_i = \log \xi_i$), $i \neq j$. The quantity $(u_i - u_j)$ represents the bias in estimating β_k that is ignored if β_k was interpreted as a PA parameter. Therefore, we need to estimate ξ_i in addition to the β 's. To do this we assume that ξ_i are independent variables drawn from a gamma family of distributions $g(\xi_i)$ with $E(\xi_i) = 1$ and $\text{Var}(\xi_i) = \theta$ (see Vaupel et. al., 1979; and Clayton, 1978).

$$g(\xi_i) = \frac{1}{\Gamma(1/\theta)} \theta^{-1/\theta} \xi_i^{1/\theta - 1} e^{-\xi_i/\theta} \quad 5.1$$

For convenience, we consider the distribution of $u_i = \ln \xi_i$, given by

$$g(u_i) = \frac{1}{\Gamma\left(\frac{1}{\theta}\right)} \theta^{-1/\theta} \exp(u/\theta - e^{u/\theta}) \quad 5.2$$

The maximum hierarchical likelihood estimates (MHLEs) are obtained by maximizing the sum of two components log-likelihood:

- (a) l_1 the log-likelihood of the data given u
- (b) l_2 the log-likelihood for u , and
- (c) $h = l_1 + l_2$

Thus h represents the logarithm of the of the hierarchical likelihood based on the joint distribution

Applied Statistics in Agriculture

of the data and u.

$$e^{\ell_1} = \prod_{i=1}^k \prod_{j=1}^{n_i} \left[(\lambda(t_{ij}|\xi_i))^{\delta_{ij}} S(t_{ij}|\xi_i) \right] \quad 5.3$$

or

$$\ell_1 = \sum_{i=1}^k \sum_{j=1}^{n_i} \left\{ \delta_{ij} \ln \lambda(t_{ij}) + \delta_{ij} u_i + \delta_{ij} x_{ij}' \beta - \Lambda_o(t_{ij}) \exp(x_{ij}' \beta + u_i) \right\} \quad 5.4$$

and

$$\ell_2 = \sum_{i=1}^k \ln g(u_i) = \sum_{i=1}^k \left\{ \frac{u_i}{\theta} - \frac{e^{u_i}}{\theta} - \frac{1}{\theta} \ln \theta - \ln \Gamma\left(\frac{1}{\theta}\right) \right\} \quad 5.5$$

The MHLE of u_i is obtained by solving $\frac{\partial h}{\partial u_i} = 0$

The MHLE of β is obtained by solving

$$\frac{\partial h}{\partial \beta_r} = \frac{\partial \ell_1}{\partial \beta_r} = 0$$

The Newton-Raphson iterative procedure for estimating β and u is

$$\begin{pmatrix} \tilde{\beta} \\ \tilde{u} \end{pmatrix} = \begin{pmatrix} \beta_o \\ u_o \end{pmatrix} + I^{-1} \begin{pmatrix} \frac{\partial \ell_1}{\partial \beta_o} \\ \frac{\partial \ell_1}{\partial u_o} \end{pmatrix} - I^{-1} \begin{pmatrix} o \\ (e^{u_i} - 1) \theta^{-1} \end{pmatrix} \quad 5.6$$

where I is the Hessian matrix.

II- Random effects model (RAEM)

The marginal likelihood function is obtained by substituting 5.1 into 3.3 and performing the integration. The log-likelihood function is written as

$$\ell(\beta, \theta) = \sum_{i=1}^k \eta_i + \sum_{i=1}^k \ln P_{\delta_i} = \ell_1^* + \ell_2^* \quad 5.7$$

where

$$\ell_1^* = \sum_{i=1}^k \sum_{j=1}^{n_i} \delta_{ij} \left(x_{ij}^t \beta + \ln \lambda_o(t_{ij}) \right)$$

and

$$\ell_2^* = \sum_{i=1}^k \left[\sum_{r=0}^{\delta_i - 1} \ln(1 + \theta r) + \delta_i \ln \mu_i - \left(\delta_i + \frac{1}{\theta} \right) \ln(1 + \theta \mu_i) \right]$$

Parameter estimates are obtained by maximizing the log-likelihood function $\ell(\beta, \theta)$, using the Newton-Raphson's iteration.

6. DATA ANALYSIS

Before fitting the models, we evaluated the test statistic to detect intraherd clustering using equations 3.4 and 3.5. We found $Z=32.94$ ($P < .0001$) and the hypothesis $H_0: \theta = 0$ was not supported by the data. Therefore, any model fitting must account for the frailty effect.

1) Population Averaged Models

a) The GJE Approach

With the GJE approach, the hazard decreases with parity and increases when mastitis is present (Table 1). The direction (sign) of the estimate is easily interpreted, however, the interpretation of the coefficients themselves is quite difficult. For example $\exp(.108) = 1.10$, has the conventional interpretation of being the constant ratio of hazards for culling between a population whose farms have the disease and another population of disease free farms. The statistical complication of the "Population Averaged" models is that each farm in the sample was a member of both populations due to the observational nature of the study design. For a detailed discussion on the GJE approach, we refer the reader to the paper by Therneau (1993).

b) The GEE Approach

As was reported by Segal and Neuhaus (1997), the GEE did not converge. We therefore used an ad-hoc approach proposed by Scott and Holt (1982). We first fit a Weibull regression model under independence and then inflated the estimated variance to obtain correlation corrected variances as follows:

$$Var(\hat{\beta}) = Var(\hat{\beta}_{indep})[1 + (n-1)\rho_x * \rho_y] \quad 6.1$$

where ρ_x and ρ_y are the intraclass correlations of the covariate x and the response variable y , and n is the average cluster size. These correlations are estimated from the one-way ANOVA set up, as described by Donner and Koval (1980).

The results of this approach are shown in Table 2. The PA models are simple to apply because the within cluster correlation is treated as a nuisance parameter, and no correct specification is required. Moreover, they can be fitted by statistical software such as SAS, and S-Plus. However, valid interpretation of the parameter estimates requires specific features for the study design.

2) Cluster Specific Approach

The results of the cluster specific model for the culling data are shown in Table 3. The cluster specific model is characterized by specifying a constant risk ratio conditional on the cluster-specific frailty. But the model attempts to estimate many parameters. In addition to the regression parameters (β 's) and the variance component θ , the unobservable random effects $\xi_1 \dots \xi_k$ are predicted as well. This means that increasing the number of sampled clusters would increase the number of quantities to be estimated. Therefore, it is not uncommon for convergence problems to occur in some situations.

It can be shown that the fitting algorithm described for this model is equivalent to the SAS IML macro provided by Schall (1991), in which the censoring indicator is treated as the dependent variable (similar to the GEE approach), while the cumulative baseline hazard is treated as an offset.

3) Random Effects Approach

Table 4 shows the results of the random effect model for the culling data. The coefficients differ in sign from the GEE because the random effects and the Weibull model fit log survival times, while the GJE and CS models fit the Cox hazard function.

7. DISCUSSION

There are fundamental differences among the models, and we must be careful when we attempt to interpret and extend their use to the analysis of data arising from observational studies. First, we realize that the proportional hazard model has been designed mainly for independent observations. The PA models treated the within cluster dependence of failure times as a nuisance factor. While ignoring the dependence may have little effect on the magnitude of the true estimated coefficients, it is well-known that their standard errors will be under estimated. To correct for the within cluster dependence we used the robust “sandwich estimator” of $\text{Var}(\hat{\beta})$.

In contrast to the PA models, the CS and the RAEM attempt to model the intra cluster correlation. However, the β coefficients need to be interpreted conditional on the unobserved frailty. This would be more naturally related to the underlying design that generated the data. The CS and the RAEM are expected to be more efficient, provided that the frailty distribution has been correctly specified.

The program codes for the GEE, GJE, Schall’s macro, and the Fortran program used for the random effects model are available from the first author (mshoukri@hotmail.com).

ACKNOWLEDGEMENTS

The first author acknowledges the financial support of the National Sciences and Engineering Research Council (NSERC) of Canada. The authors thank Ontario Dairy Herd

Improvement for assisting with data collection and the Dairy Farmers of Ontario for financial support of the second author.

REFERENCES

- Clayton, D.G. and Cuzick, J. (1985) Multivariate generalizations of the proportional hazard model (with discussion). *J. Royal Stat. Soc., Series A*, 148, 82-117.
- Clayton, D.G. (1978) A model for association in bivariate survival data and its application in epidemiologic studies of chronic disease incidence. *Biometrika*, 65, 141-151.
- Donner, A., and Koval, J. (1980) The estimation of the intraclass correlation in the analysis of family data. *Biometrics*, 36, 19-25.
- Fetrow, J. (1988) Culling dairy cows. *Proceedings of the American Association of Bovine Practitioners*, 20, 102-107.
- Greer, R.C., Whitman, R.W., and Woodward, D.R. (1980) Estimation of probability of beef cows being culled and calculation of expected herd life. *J. Anim. Sci.*, 51, 10-19.
- Huber, P.J. (1967) The behaviour of the maximum likelihood estimates under non-standard conditions. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 221-233.
- Lee, E.W., Wei, L.J. and Amato, D. (1992) Cox-type regression analysis for a large number of small groups of correlated failure time observations. In Klein, J.P. and Goel, P.K. (eds), *Survival Analysis, State of the Art*, 237-247. K Luwer Academic Publishers, Netherlands.
- Liang, K-Y. and Zeger, S.L. (1986) Longitudinal data analysis using generalized linear models. *Biometrika*, 73, 13-22.
- Lin, D.Y. and Wei, L.J. (1989) The robust inference for the Cox proportional hazards model. *J. Amer. Statist. Assoc.*, 84, 1074-1078.
- Sargeant, J.M. (1996) Factors associated with milk protein production at the provincial, herd, and individual cow level. Dissertation, University of Guelph, Guelph, Canada.
- Schall, R. (1991) Estimation in generalized models with random effects. *Biometrika*, 78, 719-727.
- Scott, A.J. and Holt, D. (1982) The effect of two-stage sampling on ordinary least squares methods. *JASA*, 77, 485-497.

Segal, M. and Neuhaus, J. (1993) Robust inference for multivariate survival data. *Statistics in Medicine*, 2, 1019-1031.

Segal, M. And Neuhaus, J. (1997) Dependence estimation for marginal models of multivariate survival data. *Lifetime Data Analysis*, 3, 251-268.

Therneau, T. (1993) Using multiple-events Cox model. *Proceedings from the Biometrics Section of the American Statistical Association*. pp. 1-14.

Vaupel, J.W., Manton, K.G. and Stallard, E. (1979) The impact of heterogeneity in individual frailty and the dynamics of mortality. *Demography*, 16, 439-454.

Table 1: The GJE approach to modelling age-at-culling.

Parameter	Estimate	SE (indep.)	SE (GJE)	Correct P-value
Parity	-1.546	0.043	.173	0.000
Milk	-.0193	0.025	.071	0.810
Disease	.108	0.067	0.065	0.107

Table 2: The Weibull model for modelling age-at-culling data.

Parameter	Estimate	SE (indep.)	SE (inflated)
Intercept	.518	.0049	.0068
Parity	.208	.0054	.0056
Milk	.003	.0095	.0108
Disease	-.01	.009	.0093

Table 3: The cluster-specific approach to modelling age-at-culling modelling via hierarchical likelihood.

Parameter	Estimate	SE
Parity	-1.015	.038
Milk	-.107	.028
Disease	.447	.054
θ	0.909	

Table 4: The random effect approach to modelling age-at-culling data.

Parameter	Estimate	SE
Parity	0.383	.026
Milk	0.092	.014
Disease	-0.009	.058
θ	.51	0.259