Journal of Applied Statistics
Vol. 38, No. 3, March 2011, 619–629

Taylor & Francis
Taylor & Francis Group

A frailty modeling approach for parental effects in animal breeding

Suely Ruiz Giolo[a]* and Clarice Garcia Borges Demétrio[b]

[a]*Department of Statistics, Federal University of Parana, Curitiba, Brazil;* [b]*Department of Exact Science, University of Sao Paulo, Piracicaba, Brazil*

(*Received 14 January 2009; final version received 26 November 2009*)

Survival models involving frailties are commonly applied in studies where correlated event time data arise due to natural or artificial clustering. In this paper we present an application of such models in the animal breeding field. Specifically, a mixed survival model with a multivariate correlated frailty term is proposed for the analysis of data from over 3611 Brazilian Nellore cattle. The primary aim is to evaluate parental genetic effects on the trait length in days that their progeny need to gain a commercially specified standard weight gain. This trait is not measured directly but can be estimated from growth data. Results point to the importance of genetic effects and suggest that these models constitute a valuable data analysis tool for beef cattle breeding.

Keywords: correlated times; frailty; growth curve; random effects; survival

1. Introduction

Frailty was first introduced in survival analysis by Vaupel *et al.* [35] in order to allow for unobserved heterogeneity. In the univariate framework, frailty models are usually taken as an extension of the proportional hazards model [5] in which both a frailty term and covariate effects are assumed to act multiplicatively on the baseline hazard. The term including covariates allows for observed heterogeneity, while the frailty term captures that part of the individual heterogeneity that refers to unobserved risk factors. In multivariate or clustered survival data, the first approach developed is based on the concept of shared frailty [4,17,29,36] in which a common random effect term acts on the hazards of all individuals in a cluster. Shared frailty models are useful when it is desired to explain correlations within groups or clusters of individuals (e.g. family, litter, clinic or recurrent events from the same individual), but they have some limitations. All unobserved risk factors are assumed to be the same within a cluster, which is not always reasonable, as more

*Corresponding author. Email: giolo@ufpr.br

ISSN 0266-4763 print/ISSN 1360-0532 online

DOI: 10.1080/02664760903521492
http://www.informaworld.com

realistically we might expect individual-level heterogeneity also. These models have also not been designed to incorporate complex genetic relationships found in family data of varying size and structure.

To overcome these limitations, a correlated frailty approach has been developed and used in several studies with bivariate survival data as, for instance, in twin studies (e.g. [18,19,21,22,38]). In the correlated frailty model, unobserved risk factors are not assumed to be the same in each group and frailty is allowed to be individual-specific exactly as in the univariate framework. However, unlike the univariate case, in the correlated frailty model there is no assumption of independence and one individual's frailty is allowed to be associated with the frailty of another individual who is related genetically. Ripatti and Palmgren [28] and Therneau *et al*. [34] extended correlated frailty models to survival data on n individuals. In this case, frailties are usually assumed to be random variables drawn from a multivariate normal distribution with an arbitrary covariance structure. Pankratz *et al*. [24] applied this model to investigate the aggregation of breast cancer within families. A correlated frailty model that incorporates both unobserved genetic and environmental sources of frailty is also applied by Garibotti *et al.* [16] to data from a family-based study of longevity.

Correlated frailty models have also been receiving increasing attention in animal breeding programmes since it was found by animal geneticists and breeders that survival analysis can be used for analyzing traits associated with longer productive life of livestock. After some analyses using survival methods were published in the animal breeding field (e.g. [13,30,31]), some computational tools have become available. Ducrocq and Söelkner [11,12], for instance, developed a package called *the Survival Kit* with animal breeding applications in mind. A Bayesian estimation approach is used in this package as described by Ducrocq and Casella [10]. Pankratz *et al.* [24] also implemented in the R software [27], a library called *kinship* [33] which contains the function *coxme* for modeling survival data from large pedigrees. For estimation they considered an approximate maximum likelihood approach. With the availability of such computational tools, routine genetic evaluation of sires for longevity of their daughters based on survival analysis was implemented in France, in Germany, in the Netherlands, and in other countries (e.g. [3,8,9,23,25,37]). Among others, dairy cattle breeds studied in such genetic evaluation were Holstein, Braundvieh, Normande, and Simmental. There has been little use in beef cattle breeding, an exception being Pereira *et al.* [26], who used survival analysis as a tool to investigate the possibility of increasing sexual precocity in Nellore, a breed extensively used in Brazil as beef cattle.

Survival models are therefore widely defined in dairy cattle for longevity breeding. However, a complication is that exact times to events of interest are not usually recorded. They are only known to occur within some time intervals that may overlap and vary in length. A way of dealing with this data would be to consider the time to event as interval-censored data. There are, however, methodological and computational difficulties related to this approach, particularly regarding the multidimensional integral involved in the likelihood, which makes it impractical for large data sets and complex genetic relationships as investigated here. Hence, in this work a genetic evaluation of Nellore sires based on length in days T that their progeny need to gain a specified weight gain from birth is performed by adopting a two-stage analysis. At the first stage we model the available intermittent growth data to estimate T. These estimates are then used in the second stage for the genetic evaluation using correlated frailty models. The aim is the possibility of decreasing this length in days by selection of sires so as to reduce costs. An application on Brazilian Nellore cattle forms the basis of the paper.

The structure of the paper is as follows: Section 2 contains details on the data set; Section 3 describes the mixed survival model and also the estimation procedure used for fitting the model to the data; Section 4 gives results from the analysis performed. A brief discussion ends the paper in Section 5.

2. The data

The data and pedigree information used in our work were supplied by the Brazilian GenSys Consultores Associados S.C. Ltda. Nellore is a Zebu breed from India that has become predominant in Brazil due to easy care, adaptability, and economic production under an intensive system. The records analyzed were from 3611 progeny of 24 Nellore sires and 3116 dams born during spring in a single herd between 1996 and 1997. The number of progeny per sire varied from 16 to 337 as can be seen in Table 1. The number per dam varied from 1 to 2 where 84% of them had 1 progeny.

All 3611 progeny were followed up from birth to approximately 2.5 years after they were born. In this follow-up period the weight of each progeny was taken six times, at intervals of approximately 3–5 months. The first weight was taken at birth. Table 2 shows descriptive statistics for the weights recorded over time. We see the growth rate falling as the animals mature but note the continuing increase in variability with age.

In addition to the weight, information recorded for each progeny included: sex, reproduction (natural or artificial), progeny birth year, and age of the dam at progeny birth, which varied from 2 to 16 years with mean and median equal to 5.5 and 4, respectively. Approximately 65% of the progeny were female and about 92% were generated by natural reproduction. About 46% of them were born in 1996, and 54% in 1997.

As weaned calves that have gained 160 kg from birth is a well-defined marketing unit in Brazil, we have interest in comparing sires based on length in days that their progeny need to gain this commercially specified weight gain. This will be considered the response time of interest in this paper. The exact time that each progeny takes to gain 160 kg is however, not precisely known since weight is taken only periodically. Thus, to estimate response time we fitted at the first stage, a logistic growth curve model for each progeny. Under this model, the weight of progeny i at time t is expressed as

$$W_i(t) = \frac{A_i}{1 + b_i \exp(-k_i t)} + \epsilon_i, \tag{1}$$

Table 1. Number of progeny generated per each one of the 24 Nellore sires.

Sire	Progeny	Sire	Progeny	Sire	Progeny	Sire	Progeny	Sire	Progeny
1	79	6	296	11	37	16	142	21	61
2	43	7	159	12	235	17	153	22	31
3	99	8	220	13	200	18	135	23	36
4	16	9	165	14	252	19	251	24	27
5	260	10	208	15	169	20	337		

Table 2. Descriptive statistics for weights recorded over time on 3611 progeny.

	Weight (kg)				
Progeny age	Mean	Minimum	Maximum	Median	Standard deviation
Birth	30.36	18	46	30	3.31
3–5 months	83.12	43	157	82	13.32
5–9 months	172.00	100	243	171	22.04
10–13 months	293.20	202	474	290	34.13
14–22 months	360.00	270	543	357	34.91
24–30 months	421.20	346	634	417	35.19

where A_i is the asymptotic weight for progeny i, commonly interpreted as the mature weight; the parameter b_i is a constant that adjusts for situations where $W(0)$ are not equal to 0; k_i is a function of the ratio of maximum growth rate to mature size, commonly referred to as maturing index [14]; and ϵ_i are random errors assumed independent and normally distributed with mean zero. To take into account the heterogeneity of the variance throughout the growth curve suggested from Table 2, we considered the variance following the same nonlinear function employed to describe the data, as proposed by Blasco *et al.* [1]. Similarly to these authors, we performed some exploratory analyses (not shown) and then concluded that the evolution of the variance could be represented following a logistic law, that is,

$$\sigma_e^2(t) = \frac{a_0}{1 + b_0 \exp(-k_0 t)}. \tag{2}$$

The overall goodness of fit of model (1) was evaluated by the correlation coefficient between the observed and predicted values.

From Equation (1), the response times t_i ($i = 1, \ldots, n$) to gain 160 kg can be then obtained by

$$t_i = \frac{\log(w_i \times \hat{b}_i) - \log(\hat{A}_i - w_i)}{\hat{k}_i}, \quad \hat{A}_i > w_i, \tag{3}$$

where w_i is the weight at birth for progeny $i + 160$ kg and $n = 3611$. Figure 1 displays a histogram of the time estimates obtained by using a Bayesian approach [32]. Priors assumed for θ_{ik} ($k = 1, 2, 3$) where $\theta_{i1} = \log(A_i)$, $\theta_{i2} = \log(b_i)$, and $\theta_{i3} = \log(-k_i)$ were $N(\mu_k, \tau_k)$ with $\mu_k \sim N(0, 10^{-4})$ and $\tau_k \sim \Gamma(10^{-3}, 10^{-3})$, where $\sigma_k^2 = 1/\tau_k$ and $\Gamma(k, \lambda)$ denotes a gamma distribution with mean k/α and variance k/α^2. For the variance (2) priors for $\theta_1 = \log(a_0)$, $\theta_2 = \log(b_0)$, and $\theta_3 = \log(-k_0)$ were assumed to be similar to those reported for θ_{ik} ($k = 1, 2, 3$). The posterior mean of the parameters were obtained by carrying out a chain of 20,000 iterations after discarding the first 5000.

A correlation coefficient between the observed and predicted values of 0.954 suggested a satisfactory overall goodness of fit of model (1). Time estimates shown in Figure 1 will be then considered as the response of interest in the survival model presented next. Note that no progeny is expected to gain 160 kg before 167 days. Also, there are no censored data since all progeny have gained at least 160 kg in the follow-up period (Table 2). Table 3 shows the first five lines of the Nellore cattle data available for the second stage of the analysis.

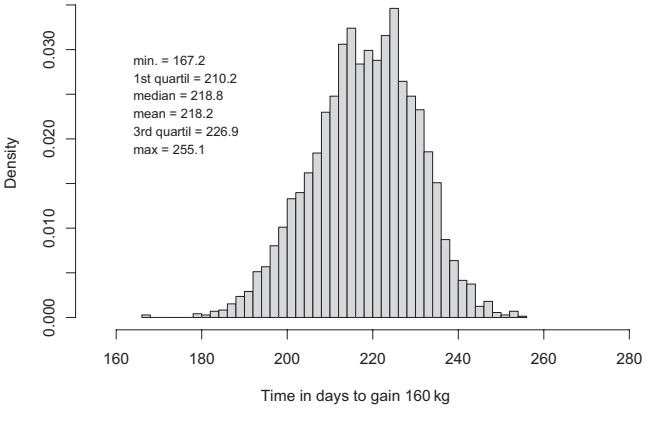

Figure 1. Histogram of the time to gain 160 kg estimated from logistic curves.

Table 3. First five lines of the Nellore cattle data where pby is the progeny birth year, agd is the age of the dam at progeny birth, and t_i is the posterior mean of the time to gain 160 kg obtained from logistic growth curves.

Progeny	Sire	Dam	Sex	pby	Agd	t_i	95% confidence interval of t_i
1	1	1	Male	1997	11	224.0	(208.3, 241.3)
2	1	2	Female	1997	10	221.8	(206.8, 238.2)
3	1	3	Male	1997	9	214.0	(200.1, 229.8)
4	1	4	Male	1997	12	205.5	(192.5, 220.1)
5	1	5	Female	1996	6	234.5	(218.3, 252.6)

3. Survival mixed model

We now outline the correlated frailty model to be applied to the Brazilian Nellore cattle data described in Section 2. Letting $\mathbf{x}_{ij}$ be a vector of measured covariates, we assume that the hazard function at time t of an animal i ($i = 1, \ldots, n_j$), which is progeny of sire j ($j = 1, \ldots, q$) can be expressed as

$$\lambda_{ij}(t) = \lambda_0(t) \exp\{\boldsymbol{\beta}'\mathbf{x}_{ij} + \omega_{ij}\}, \tag{4}$$

where $\lambda_0(t)$ is the baseline hazard function, which can be left completely unspecified (Cox model) or may follow a parametric distribution (e.g. Weibull), $\boldsymbol{\beta}$ is a vector of unknown regression coefficients, and $\boldsymbol{\omega} = \{\omega_{ij}\}$ is a vector of unobserved frailties with ω_{ij} representing the per-progeny random effect.

Ducrocq and Casella [10] describe how it is usual in quantitative genetics to assume the vector of frailties $\boldsymbol{\omega}$ follows a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Omega}$. To evaluate the influences of unobserved genetic contribution to time to gain 160 kg, this covariance matrix is given by $\boldsymbol{\Omega} = 2\boldsymbol{\Phi}\sigma_g^2$, where σ_g^2 represents the shared polygenic effect influences. The block diagonal matrix $2\boldsymbol{\Phi}$, where each block has dimensions $n_j \times n_j$, is the so-called relationship matrix that captures the shared polygenic factors between genetically related family members. Element $\phi_{ii'}$ of $2\boldsymbol{\Phi}$ expresses the degree of genetic resemblance between pair (i, i') of progeny from sire j, i.e. the elements $\{\phi_{ii'}\}$ represent the expected proportion of the genome that is shared by each pair of progeny. The elements $\{\phi_{ii'}\}$ take values 1 on the diagonal, 0 for pairs which are not genetically related, 0.5 for first-degree relatives (e.g. sibling pairs), 0.25 for second-degree relatives (e.g. uncle/nephew), 0.125 for third-degree relatives (e.g. first cousins), and so forth. Thus, for $i, i' = 1, \ldots, n_j$,

$$\phi_{ii'} = \begin{cases} 0 & \text{if } i \text{ not genetically related with } i', \\ 1 & \text{if } i = i', \\ \left(\frac{1}{2}\right)^r & \text{if } i \neq i' \text{ and genetically related, } r = 1, 2, \ldots \end{cases}$$

with r denoting the degree of relationship between progeny i and i'.

In order to evaluate simultaneously the influences of unobserved genetic and shared family environmental contributions to time to gain 160 kg, the covariance matrix $\boldsymbol{\Omega}$ can be decomposed into two components, $\boldsymbol{\Omega} = 2\boldsymbol{\Phi}\sigma_g^2 + \boldsymbol{\Psi}\sigma_f^2$, where σ_g^2 and σ_f^2 represent the shared polygenic effect and the shared family environmental influences, respectively. Here, $\boldsymbol{\Psi}$ is also a block diagonal matrix that incorporates the degree of shared environment among progeny of sire j in which each block is a matrix of dimensions $n_j \times n_j$. If, for instance, there is an indication of similar environmental influences shared among progeny of sire j, $\boldsymbol{\Psi}$ can be assumed as a block diagonal matrix in which each block is a matrix whose elements are all equal to 1. For other alternatives for $\boldsymbol{\Psi}$ see, for instance, Yip *et al.* [39].

The conditional survival function for model (4) is expressed as

$$S_{ij}(t|\omega_{ij}) = \exp\left\{-\int_0^t \lambda_0(v)\exp\{\boldsymbol{\beta}'\mathbf{x}_{ij} + \omega_{ij}\}\mathrm{d}v\right\}$$
$$= [S_0(t)]^{\exp\{\boldsymbol{\beta}'\mathbf{x}_{ij} + \omega_{ij}\}}, \tag{5}$$

where $S_0(t) = \exp\{-\int_0^t \lambda_0(v)\,\mathrm{d}v\}$ is the baseline survival function. We also define $H_0(t) = \int_0^t \lambda_0(v)\mathrm{d}v$, the cumulative baseline hazard function usually estimated by the Breslow's estimator [2]. Measured covariates considered in the model were: progeny sex, progeny birth year, and age of the dam at progeny birth.

Assuming that the random effects of all progeny of all the sires follow a multivariate normal distribution with mean zero and covariance matrix $\boldsymbol{\Omega}$, maximization of the partial likelihood (PL) given by

$$L = \int \mathrm{PL}(\boldsymbol{\beta}, \mathbf{g})\frac{1}{\sqrt{2\pi|\boldsymbol{\Omega}|}}\exp\left\{-\frac{1}{2}\,\mathbf{g}'\boldsymbol{\Omega}^{-1}\mathbf{g}\right\}\mathrm{d}\mathbf{g}, \tag{6}$$

where PL is the Cox partial likelihood [6], is performed as described by Pankratz *et al.* [24] who used Laplace approximation as suggested by Ripatti and Palmgren [28] in order to overcome the intractable multidimensional integral in Equation (6).

From the breeder's perspective, functions of the genetic parameter σ_g^2 are of particular interest, especially the *heritability*. This measures the proportion, on a suitable scale, of the total variability of the trait caused by genetic differences among the animals on which the measurements were taken. Although for the mixed-effects Cox model (4) it is not possible to obtain direct heritability estimates as in the variance components model, since there is no random error variance component, the polygenic variance component obtained from model (4) may be interpreted as measures of familial aggregation. Information concerning hazard ratios associated with the disease that corresponds to the random effect is obtained by exponentiation of the square root of the polygenic variance component. Hazard ratios associated with the covariates can also be obtained by exponentiation of each regression coefficient.

Diagnostic methods for assessing the fit of model (4) are not quite easy to define given that the definition of residual is not as clear-cut for this model. Additional studies are therefore needed for this purpose. However, as the validity of model (4) relies heavily on the assumption of proportional hazards, we used a graphical technique [20] as an exploratory method to investigate the acceptability of this assumption. Based on this technique, each covariate is stratified into k disjoint strata. Stratified Cox models, in which a distinct baseline hazard function $\lambda_{0j}(t)$ is assumed for each stratum, are then fitted for each covariate to obtain the estimated cumulative baseline hazard for each stratum j, that is, $\hat{H}_{0j}(t)$, $j = 1, \ldots, k$. Indications that the assumption holds are suggested if for each covariate the log cumulative baseline hazards $\ln[\hat{H}_{01}(t)], \ldots, \ln[\hat{H}_{0k}(t)]$ plotted against t show no gross departure of parallel curves.

4. Results and discussion

Table 4 shows the main parameters estimated from Cox mixed survival model (4) considering three covariance matrix. First, it was taken $\boldsymbol{\Omega} = \boldsymbol{\Psi}\sigma_f^2$ for evaluating only the influences of unobserved shared family environmental contributions to time to gain 160 kg. The second covariance matrix considered was $\boldsymbol{\Omega} = 2\boldsymbol{\Phi}\sigma_g^2$ which allows us to evaluate only the influences of unobserved genetic contribution. Finally, it was taken $\boldsymbol{\Omega} = 2\boldsymbol{\Phi}\sigma_g^2 + \boldsymbol{\Psi}\sigma_f^2$ for evaluating both of these influences simultaneously. Since all progeny share the same climate, pasture and other factors, it was assumed that progeny of the same sire shared similar environmental influences. Hence, all elements of each

Table 4. Parameter estimates and their respective 95% confidence intervals where pby is the progeny birth year and agd is the age of the dam at progeny birth.

Matrix $\mathbf{\Omega}$	Coefficient estimates				Variance components	
	Sex	pby	Agd	Agd^2	σ_f^2	σ_g^2
$\mathbf{1}\sigma_f^2$	1.82 (1.73, 1.90)	0.28 (0.08, 0.48)	0.06 (0.00, 0.12)	−0.003 (−0.007, 0.001)	0.077 (0.01, 0.20)	–
$2\mathbf{\Phi}\sigma_g^2$	2.18 (2.08, 2.28)	0.30 (0.12, 0.47)	0.07 (0.06, 0.10)	−0.004 (−0.007, −0.001)	–	0.368 (0.12, 0.81)
$2\mathbf{\Phi}\sigma_g^2 + \mathbf{1}\sigma_f^2$	2.24 (2.13, 2.34)	0.30 (0.19, 0.40)	0.07 (0.03, 0.11)	−0.004 (−0.008, −0.001)	0.017 (0.0, 0.03)	0.458 (0.16, 0.92)

block of the matrix $\mathbf{\Psi}$ were assumed to be equal to 1. Covariates included in the model were sex, progeny birth year, and age of the dam at progeny birth. As progeny from youngest and also oldest dams are expected to have slower growth rate, age of the dam at progeny birth was considered as having a second-order polynomial effect. Estimates of the variance components σ_f^2 and σ_g^2 are therefore adjusted for such terms in the model. Because it is very hard to obtain the standard errors (SE) for the variance components, confidence intervals are obtained by profiling the likelihood [7]. All computations were done by using the library kinship [27,33].

As mentioned in Section 3, hazard ratios can be obtained by exponentiation of each regression coefficient. Hence, from model which, for instance, incorporated both polygenic and environmental effects, we have for the covariate sex that $\exp(2.24) = 9.39$ with 95% confidence interval (CI) of (8.41, 10.38). Therefore, the hazard of obtaining the desirable gain of weight in a shorter period of time is estimated to be much greater among males than among females progeny. Evidence of a sex effect is then observed with faster growth for males. Similarly, the hazard of obtaining the desirable gain of weight in a shorter period of time is estimated to be 35% ($\text{CI}_{95\%}$, 21−49%) greater among progeny births in 1997 than among those births in 1996. Note that the proportional hazards assumption is suggested from Figure 2 since no gross departure of parallel curves are observed for the covariates in the survival model.

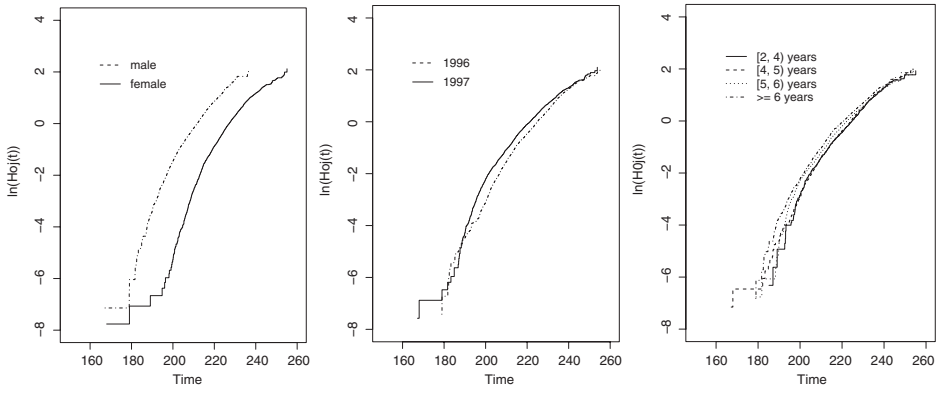

Figure 2. Log cumulative baseline hazards $\ln(\hat{H}_{0j}(t))$, $j = 1, \ldots, k$, versus t for the covariates (a) sex into $k = 2$ strata, (b) progeny birth year into $k = 2$ strata, and (c) age of the dam at progeny birth into $k = 4$ strata ([2, 4), [4, 5), [5, 6) and ≥ 6 years).

Information regarding hazard ratios associated with the random effects, in which our main interests are focused, can also be obtained by exponentiation of the square root of the variance components. From estimates shown in Table 4 for these components, it can be observed higher estimates associated with the polygenic effect σ_g^2. There is, therefore, evidence that polygenic factors play a greater role than shared family environment factors in explaining the variation in the time to gain the desirable weight. The estimate of the polygenic variance component equal to 0.458 ($CI_{95\%}$, $0.16 - 0.92$) suggests a significant degree of heritability associated with the time to gain 160 kg and shows that the progeny hazard of obtaining this weight gain in a shorter period of time due to polygenic effects are on average $\exp(\sqrt{0.458}) \approx 1.967$. Hence, there are progeny which have hazard of gaining 160 kg in a shorter period of time 96.7% higher than the overall average hazard for the entire sample. From this result we have therefore evidence of large genetic effects in growth rate of progeny.

In the animal breeding field the random effect ω_{ij} is referred to as the genetic value of the progeny i from sire j. Figure 3 shows these progeny genetic values per sire obtained from the model which includes both polygenic and environmental random effects. Sires with mean genetic value of their progeny significantly greater than zero can be identified from the figure. For example, progeny from sires 2, 9, 10, 19, and 24 seem to have a proven performance than progeny from other sires regarding the time to reach the desirable gain of weight. Further selectivity in sires may provide therefore the potential to decrease the length in days to gain the required weight.

For illustration purposes we take two sires amongst the 24 in the study, sire 10 with high progeny genetic mean value (0.89), and sire 4 with low progeny genetic mean value (-1.07). Estimated mean survival curves, expressed in Equation (5), for progeny of these two sires are displayed in Figure 4. Note from the male survival curves at time $t = 190$ days (vertical line in the plots) that $\hat{S}(190) \approx 0.009$ for sire 10 and 0.516 for sire 4. This means that only about 1% of male progeny from sire 10 have *not* gained the desired weight at such time as against 52% from sire 4. Corresponding values for female progeny, with slower growth rate, are 60.6% and 93.2%. Thus the effect of high and low genetic values is clearly observed.

If we consider that one objective in cattle breeding programmes is to develop methods of identifying sires with best performance based on animals generated by them, sire selection based on their progeny genetic mean values could help breeders to more quickly produce animals with the required weight for the market.

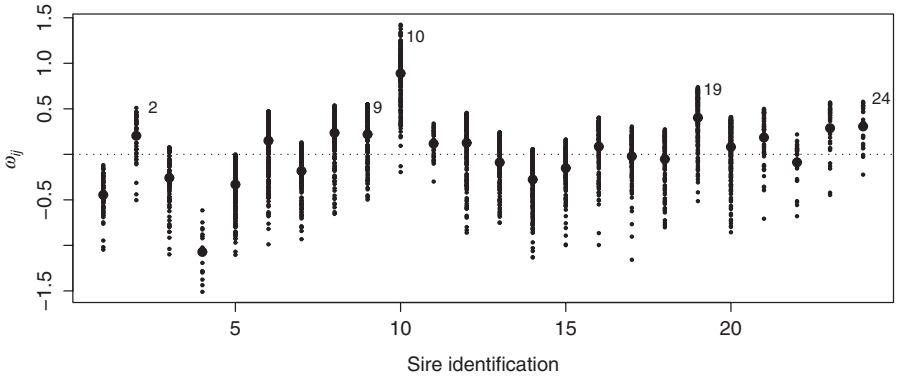

Figure 3. Progeny genetic values ω_{ij} per sire obtained from mixed Cox model with polygenic and environmental random effects, and mean genetic value by sire.

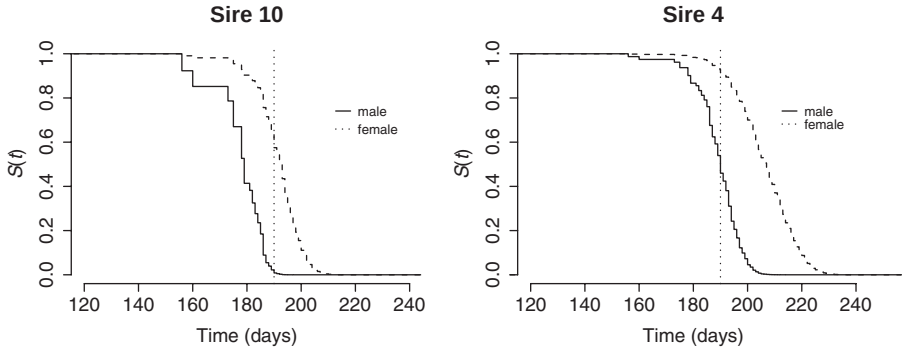

Figure 4. Mean survival curves estimated from Cox model for progeny from sires 4 and 10 by sex, progeny birth year 1996 (reference category), and mean age of the dam at progeny birth (5.5 yrs).

5. Conclusions and remarks

In this paper we have used mixed survival models to investigate the possibility of decreasing the time to event, i.e. the length in days that progeny take to gain 160 kg from birth, by selection of sires. Since in practice measures of weight cannot be made daily, it follows that this exact time is commonly unknown. For example, in the application considered here, the weight of progeny was taken six times. A way of dealing only with this routinely recorded data would be to consider the time to event as interval-censored data. As there are, however, methodological and computational difficulties related to this approach, we have first estimated the response time by considering the logistic growth curve, which has been shown to be appropriate to describe Nellore cattle growth [15], and then we fitted a correlated frailty model designed to incorporate genetic relationship between progeny. In this model we assumed the vector of random effects following a multivariate normal distribution. Because in some cases this assumption might not be reasonable, some additional efforts remain needed and would be useful to extend this model for other distributions.

Overall, our results suggest that use of statistical analyses which consider survival traits and which can have a significant impact on costs should be encouraged in the animal breeding field. In this direction, survival models can be seen as a valuable tool that can help breeders to evaluate and subsequently improve their system of production. From the Nellore cattle data analyzed in this work, results indicate that the length in days that progeny take to gain a specified weight gain from birth can be decreased by selection of sires. Sires in which the mean genetic value of their progeny are higher are desirable in the sense that they can generate progeny reaching the desired weight gain in shorter periods of time with consequent reduced costs and a possibly significant impact on profitability.

Acknowledgements

We thank the GenSys for providing the data, Robin Henderson and two anonymous referees for their valuable comments and suggestions, and the Brazilian agency CAPES for the partial financial support given to the first author (Grant BEX 0298/01-8).

References

[1] A. Blasco, M. Piles, and L. Varona, *A Bayesian analysis of the effect of selection for growth rate on growth curves in rabbits*, Genet. Sel. Evol. 35 (2003), pp. 21–41.

[2] N.E. Breslow, *Contribution to the discussion on the paper by D.R. Cox*, J. Roy. Statist. Soc. Ser. B 34 (1972), pp. 216–217.
[3] A. Buenger, V. Ducrocq, and H.H. Swalve, *Analysis of survival in dairy cows with supplementary data on type scores and housing systems from a region of northwest germany*, J. Dairy Sci. 84 (2001), pp. 1531–1541.
[4] D.G. Clayton, *A model for association in bivariate life tables and its application in epidemiological studies of family tendency in chronic disease incidence*, Biometrika 65 (1978), pp. 141–151.
[5] D.R. Cox, *Regression models and life tables (with discussion)*, J. Roy. Statist. Soc. Ser. B 34 (1972), pp. 187–220.
[6] D.R. Cox, *Partial likelihood*, Biometrika 62 (1975), pp. 269–276.
[7] M. de Andrade, E.J. Atkinson, E. Lunde, C.I. Amos, and J. Chen, *Estimating genetic Components of Variance for Quantitative Traits in Family studies using the MULTIC routines*, Tech. Rep. 78, Mayo Foundation, Rochester, MN, USA, 2006. Available at http://mayoresearch.mayo.edu/mayo/research/biostat/upload/78.pdf.
[8] G. de Jong, A.R. Vollema, S. van der Beek, and A. Harbers, *Breeding Value for Functional Longevity in the Netherlands*, Proceedings of the International Workshop on EU Concerted Action Genetic Improvement of Functional Traits in Cattle (GIFT), Vol. 21, Jouy-en-Josas, France, 1999, pp. 68–72.
[9] V. Ducrocq, *Statistical analysis of length of productive life of dairy cows of the Normande breed*, J. Dairy Sci. 77 (1994), pp. 855–866.
[10] V. Ducrocq and G. Casella, *A Bayesian analysis of mixed survival models*, Genet. Sel. Evol. 28 (1996), pp. 505–529.
[11] V. Ducrocq and V.J. Sölkner, *The Survival Kit, a Fortran Package for the Analysis of Survival Data*, Proceedings of the 5th World Congress on Genetics Applied to Livestock Production Vol, 21, University of Guelph, Ontario, Canada, 1994, pp. 51–52.
[12] V. Ducrocq and V.J. Sölkner, *The Survival Kit-V3.0: a Package for Large Analyses of Survival Data*, Proceedings of the 6th World Congress on Genetics Applied to Livestock Production, Vol. 27, Armidale, New South Wales, Australia, 1998, pp. 447–448.
[13] T.R. Famula, *Exponential stayability model with censoring and covariances*, J. Diary Sci. 64 (1981), pp. 538–545.
[14] H.A. Fitzhugh, Jr, *Analysis of growth curves and strategies for altering their shape*, J. Anim. Sci. 42 (1976), pp. 1036–1051.
[15] A.R. Freitas, *Growth curves in animal production*, Revista Brasileira de Zootecnia 34 (2005), pp. 786–795.
[16] G. Garibotti, K.R. Smith, R.A. Kerber, and K.M. Boucher, *Longevity and correlated frailty in multigenerational families*, J. Gerentol. A Biol. Sci. 61 (2006), pp. 1253–1261.
[17] P. Hougaard, *Life table methods for heterogeneous populations: distributions describing the heterogeneity*, Biometrika 71 (1984), pp. 75–83.
[18] S. Kheiri, A. Kimber, and M.R. Meshkani, *Bayesian analysis of an inverse Gaussian correlated frailty model*, Comput. Statist. Data Anal. 51 (2007), pp. 5317–5326.
[19] S. Kheiri, M.R. Meshkani, and S. Faghihzadeh, *A correlated frailty model for analysing risk factors in bilateral corneal graft rejection for Keratoconus: a Bayesian approach*, Statist. Med. 24 (2005), pp. 2681–2693.
[20] J.P. Klein and M.L. Moeschberger, *Survival Analysis: Techniques for Censored and Truncated Data*. Springer, New York, 2003.
[21] I. Locatelli, P. Lichtenstein, and A.I. Yashin, *A Bayesian correlated frailty model applied to Swedish breast cancer data*, MPIDR Working Paper, WP2003-025, 2003. Available at http://www.demogr.mpg.de/papers/working/wp-2003-025.pdf.
[22] I. Locatelli, A. Rosina, P. Lichtenstein, and A.I. Yashin, *A correlated frailty model with long-term survivors for estimating the heritability of breast cancer*, Statist. Med. 26 (2006), pp. 3722–3734.
[23] J.M.K. Ojango, V. Ducrocq, and G.E. Pollott, *Survival analysis of factors affecting culling early in the productive life of Holstein–Friesian cattle in Kenya*, Livest. Prod. Sci. 92 (2005), pp. 317–322.
[24] V.S. Pankratz, M. de Andrade, and T.M. Therneau, *Random-effects Cox proportional hazards model: general variance components methods for time-to-event data*, Genet. Epidemiol. 28 (2005), pp. 97–109.
[25] E. Pasman and F. Reinhardt, *Genetic Evaluation for Length of Productive Life of Holstein cattle in Germany*, Proceedings of the International Workshop on EU Concerted Action Genetic Improvement of Functional Traits in Cattle (GIFT), Vol. 21, Jouy-en-Josas, France, 1999, pp. 55–59.
[26] E. Pereira, H.N. Oliveira, J.P. Eler, J.A. II de V. Silva, and M.H. Melis, *Use of survival analysis as a tool for the genetic improvement of age at first conception in Nellore cattle*, J. Anim. Breed. Genet. 123 (2006), pp. 64–71.
[27] R Development Core Team. *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, 2008. ISBN 3-900051-07-0. Available at http://www.R-project.org.
[28] S. Ripatti and J. Palmgren, *Estimation of multivariate frailty models using penalised partial likelihood*, Biometrics 56 (2000), pp. 1016–1022.
[29] K.S. Sahu, D.K. Dey, H. Aslanidou, and D. Sinha, *A Weibull regression model with gamma frailties for multivariate survival data*, Lifetime Data Anal. 3 (1997), pp. 123–137.
[30] S.P. Smith, *The extension of failure time analysis to problems of animal breeding*, Ph.D. diss., Cornell University, Ithaca, NY, 1983.

[31] S.P. Smith and R.L. Quaas, *Productive life span of bull progeny groups: failure time analysis*, J. Diary Sci. 67 (1984), pp. 2999–3007.
[32] D. Spiegelhalter, A. Thomas, and N. Best, *WinBugs user manual*, 2000. Available at http://www.mrc-bsu.cam.ac.uk/bugs.
[33] T.M. Therneau, *On mixed-effect Cox models, sparse matrices, and modeling data from large pedigrees*, Tech. Rep., Mayo Foundation, Rochester, MN, USA, 2007. Available at http://mayoresearch.mayo.edu/biostat/upload/kinship.pdf.
[34] T.M. Therneau, P.M. Grambsch, and V.S. Pankratz, *Penalized survival models and frailty*, J. Comput. Graph. Statist. 12 (2003), pp. 156–175.
[35] J.W. Vaupel, K.G. Manton, and E. Stallard, *The impact of heterogeneity in individual frailty on the dynamics of mortality*, Demography 16 (1979), pp. 439–454.
[36] J.W. Vaupel, A.I. Yashin, L. Xiu, N.V. Holm, and B. Harvald, *Survival Analysis in Genetics: Danish twin Data Applied to a Gerontological Question*, in Proceedings of the NATO ASI Series, Serie E: Applied Science, J.P. Klein, ed., Kluwer, The Netherlands, 1992.
[37] N. Vukasinovic, J. Moll, and L. Casanova, *Implementation of a routine genetic evaluation for longevity based on survival analysis techniques in dairy cattle populations in Switzerland*, J. Dairy Sci. 84 (2001), pp. 2073–2080.
[38] A.I. Yashin and I.A. Iachine, *Genetic analysis of durations: correlated frailty models applied to survival of Danish twins*, Genet. Epidemiol. 12 (1995), pp. 529–538.
[39] B.H. Yip, C. Bjork, P. Lichtenstein, C.M. Hultman, and Y. Pawitan, *Covariance component models for multivariate binary traits in family data analysis*, Statist. Med. 27 (2008), pp. 1086–1105.