Original article

Analysis of rabbit doe longevity using a semiparametric log-Normal animal frailty model with time-dependent covariates

Juan Pablo Sánchez[a*], Inge Riis Korsgaard[b],
Lars Holm Damgaard, Manuel Baselga[b]

[a] Departamento de Ciencia Animal, Universidad Politécnica de Valencia, Camino de Vera 14, 46022 Valencia, Spain
[b] Department of Genetics and Biotechnology, Research Centre Foulum, Danish Institute of Agricultural Sciences, P.O. Box 50, DK-8830 Tjele, Denmark

Abstract – Data on doe longevity in a rabbit population were analysed using a semiparametric log-Normal animal frailty model. Longevity was defined as the time from the first positive pregnancy test to death or culling due to pathological problems. Does culled for other reasons had right censored records of longevity. The model included time dependent covariates associated with year by season, the interaction between physiological state and the number of young born alive, and between order of positive pregnancy test and physiological state. The model also included an additive genetic effect and a residual in log frailty. Properties of marginal posterior distributions of specific parameters were inferred from a full Bayesian analysis using Gibbs sampling. All of the fully conditional posterior distributions defining a Gibbs sampler were easy to sample from, either directly or using adaptive rejection sampling. The marginal posterior mean estimates of the additive genetic variance and of the residual variance in log frailty were 0.247 and 0.690.

longevity in rabbit does / semiparametric log-Normal frailty model / survival analysis / time dependent covariates

1. INTRODUCTION

In rabbit does for meat production the average annual replacement rate is around 120% [15]. Moreover, around 50% of the dead and culled does are in one of the first three kiddings [16]. This implies that the possibilities for culling

* Corresponding author: juansan@dca.upv.es
Present address: Departamento de Producción Animal I. Universidad de León, 24071 León, Spain

due to low production are limited and that a high proportion of does are dead or culled before the investment in buying them is recovered. It is therefore desirable to increase longevity of rabbit does and one possibility to achieve this objective is through directional selection. In order to improve longevity through selective breeding, it is necessary to demonstrate that additive genetic variation exists.

In rabbits, one male is mated with a few does, the proportion of full-sibs in the data is high and females are closely related. In populations with this family structure the assumptions underlying sire models are far away from being satisfied, and to use animal models should be preferred. Another characteristic of rabbits is the very high reproductive rhythm, which implies that a female changes from one physiological state (*e.g.* lactating, pregnant, empty) to another within just a few days. The risk of culling/death is expected to differ in different physiological states, therefore it is desirable to use an animal model with time dependent covariates for analysing longevity records of rabbits.

Apart from a study of Diaz *et al.* [5] where they used the *Survival Kit* [7] to estimate the additive genetic variance for longevity of beef cattle using an animal model, we are not aware of any other applications of the *Survival Kit* in which an animal model has been applied to infer the additive genetic variance for a longevity trait. In a simulation study, Ducrocq and Casella [6] analysed data according to a balanced halfsib design with an animal model using the *Survival Kit* and found that the estimated sire component of the variance was upward biased. They argued that this bias may be due to limitations of the Laplacian integration used to obtain the marginal posterior distribution of the genetic variance component, because of the high number of parameters to integrate out in animal models and because of the lack of relationships on the female side. It is likely that these simulation results, in combination with the larger computation time in animal models compared to sire models, have led to the domination of sire models over animal models in genetic applications [9,17,18]. Another reason may be that survival models have mainly been used to address genetic aspects of longevity of dairy cows in which the family structure ressembles that of a halfsib design to a higher extent than that of a rabbit population.

The semiparametric animal frailty model used by Korsgaard *et al.* [13] to analyse records of time until the first occurrence of a respiratory disease of beef bulls is a generalisation of the corresponding model in the *Survival Kit* in that it includes a normally distributed residual effect on the log-frailty scale. This model recognises that individual heterogeneity, due to environmental factors, generally exists. Korsgaard *et al.* [13] derived and implemented the animal

model without time-dependent covariates and in a Bayesian setting using Gibbs sampling, from which the marginal posterior distribution of the additive genetic variance and the residual variance in log-frailty were easily obtained.

The purpose of this paper was to infer additive genetic and environmental aspects of the length of the productive life of rabbit does using an animal model. In order to achieve this objective, we extended the model proposed by Korsgaard *et al.* [13] to include time dependent covariates.

2. MATERIALS AND METHODS

2.1. Data

Data on rabbit doe longevity were collected from the beginning of 1992 to the end of 2001 in a selection nucleus of rabbits, located in Sant Carles de la Rápita (Tarragona, Spain). The animals belong to the V strain [8], where animals were selected for litter size, 28 days after parturition, based on a family index. In this nucleus, the does were mated 11 days postpartum, the pregnancy test was carried out by abdominal palpation 12 days after mating and the young are weaned at 35 days of age.

The data set included reproductive records of the first 13 generations, involving 2400 does. The pedigree file included a total of 3031 animals. On average, one male was mated with 2.2 females having 7.7 offspring, and one female was mated with 1.03 males having 3.6 offspring. Because of space limitations and the breeding programme of the nucleus, 50% of the does were removed before the end of their productive life – these animals had censored records of longevity. The remaining 50% of the does either died or were culled because of pathological problems – these animals had uncensored records of longevity. The animals still alive at day 586 (maximum uncensored record of longevity) were censored at age 586. Summary statistics for censored and uncensored records are presented in Table I.

The records of individual animals included the date of first positive pregnancy test and the date of death or culling. The response variables were the time between these dates and the censoring code. Individual records also included dates of the different positive pregnancy tests, the number of young born alive in every parturition, and the length of the different physiological states (Pregnant, Lactating, Empty and Pregnant&Lactating).

2.2. Methods

The semiparametric log-Normal animal frailty model [3] was used to analyse rabbit doe longevity. Model parameters were inferred from a fully

Table I. Minimum (Min), Maximum (Max), and Average (Mean) values (in days) among censored and uncensored records of rabbit doe longevity.

	Censored	Uncensored
Min	13	1
Max	586	586
Mean	377.9	188.2

Bayesian analysis using Gibbs sampling [10]. After having specified the conditional likelihood and the prior distribution, a set of fully conditional posterior distributions defining a Gibbs sampler was derived. Some of the fully conditional posterior distributions are well known distributions that are easy to sample from, the remaining can all be shown to be log-concave and Adaptive Rejection Sampling (ARS) was used to obtain samples from these distributions [12].

2.2.1. Model

Let T_i and C_i be random variables representing the survival and the censoring time of animal i. The data on animal i are (y_i, δ_i), where y_i is the observed value of $Y_i = \min\{T_i, C_i\}$ and δ_i is an indicator random variable, equal to 1 if $T_i \leq C_i$, and 0 if $T_i > C_i$. $(\mathbf{y}, \boldsymbol{\delta})$ will denote data on all animals. In the semiparametric frailty model [2] it is assumed that the hazard function of T_i conditional on random effects, $\mathbf{a}$ and $\mathbf{e}$, and model parameters $\boldsymbol{\theta} = \left(\lambda_0(\cdot), \boldsymbol{\beta}, \sigma_a^2, \sigma_e^2\right)$ is given by

$$\lambda_i(t|\boldsymbol{\theta}, \mathbf{a}, \mathbf{e}) = \lambda_0(t) \exp\left\{\mathbf{x}_i'(t)\boldsymbol{\beta} + a_i + e_i\right\}, \tag{1}$$

where $\lambda_0(t)$ is the baseline hazard function, which is approximated by a piecewise constant function given by $\lambda_0(t) = \lambda_{0m}$ for $t \in (\tau_{m-1}, \tau_m]$; $m = 1, ..., M$, where $\tau_1, \cdots, \tau_M$ are the M different ordered survival times, $0 < \tau_1 < \cdots < \tau_M < \infty$; $\tau_0 = 0$ and $\tau_{M+1} = \infty$. $\mathbf{x}_i'(t)$ is the time dependent design vector of animal i, which is assumed to be a piecewise constant function of time, and $\boldsymbol{\beta}$ is the corresponding vector of regression parameters. In this study, all of the elements of $\mathbf{x}_i'(t)$ are either zeros or ones. $\mathbf{a} = (a_i)_{i=1,...,N}$ is a vector of additive genetic values, $\mathbf{a} \mid \boldsymbol{\theta} \sim N_N(0, \mathbf{A}\sigma_a^2)$, where N is the number of animals in the pedigree, and $\mathbf{e} = (e_i)_{i=1,...,n}$ is a vector of residuals in log frailty, $\mathbf{e} \mid \boldsymbol{\theta} \sim N_n(0, \mathbf{I}_n\sigma_e^2)$, where n is the number of animals with record. $\mathbf{a}$ and $\mathbf{e}$ are assumed to be independent given $\boldsymbol{\theta}$. In this model $\exp\{a_i + e_i\}$ is the frailty term.

In this study β included effects of the following covariates:

- A time dependent Year-Season effect (YS) with 37 levels. The level changes every 3 months (1st January, 1st April, 1st July and 1st October every

year). This factor accounts for the effect acting over all contemporary animals on the farm.

- A time dependent interaction between the order of positive pregnancy test (OPPT) (1st, 2nd and $\geq$ 3rd) and physiological state (PS) (Pregnant, Lactating, Empty and Pregnant&Lactating). A doe was considered Pregnant after a positive pregnancy test, Lactating after parturition if at least one young was born alive, Empty if no young were born alive, or after end of lactation if the doe had a negative pregnancy test, and Pregnant&Lactating in the part of the lactation period where the doe was pregnant (tested positive for pregnancy). Note that, by definition, females in the first order of a positive pregnancy test can not be included in the physiological state Pregnant&Lactating.

- A time dependent interaction between PS and number born alive (BA), where BA has nine levels: The first level included does that were not nulliparous and had zero born alive, does with one or two born alive were included in the second level, does with three or four born alive were included in the third level and so on until the eighth level, which included the does with 13 or more born alive. All of the does were included in the ninth level of BA before first parturition. Note that a female having zero born alive at a particular parturition can not be included in the physiological states Lactating and Pregnant&Lactating in that kindling.

All these factors were included in the model because their effects were found to be significantly different from zero, based on the likelihood ratio tests carried out in the *Survival Kit* [7] using a Cox model. In order to ensure the identifiability of regression parameters ('fixed effects') in the model, a set of constraints were imposed.

2.2.2. *Conditional likelihood*

In the following we let $\boldsymbol{\psi} = (\boldsymbol{\theta}, \mathbf{a}, \mathbf{e})$ denote the augmented vector of parameters. Under the assumption that conditional on $\boldsymbol{\psi}$ censoring is independent and noninformative of $\boldsymbol{\psi}$ [2] the conditional likelihood can be written as

$$p(\mathbf{y}, \boldsymbol{\delta} \mid \boldsymbol{\psi}) \propto \left[\prod_{m=1}^{M} \lambda_{0m}^{d(\tau_m)} \right] \times \left[\prod_{m=1}^{M} \prod_{i \in D(\tau_m)} z_i \exp\left\{ \mathbf{x}'_i(y_i)\boldsymbol{\beta} \right\} \right]$$
$$\times \exp\left\{ -\sum_{m=1}^{M} \lambda_{0m} \left[\sum_{i \in R(\tau_m)} z_i \int_{\tau_{m-1}}^{\tau_m \wedge y_i} \exp\left\{ \mathbf{x}'_i(s)\boldsymbol{\beta} \right\} ds \right] \right\}, \quad (2)$$

where $z_i = \exp\{a_i + e_i\}$, $D(\tau_m)$ is the set of animals that fail in $(\tau_{m-1}, \tau_m]$, (*i.e.* at time τ_m), $d(\tau_m)$ is the number of animals in $D(\tau_m)$ and $R(\tau_m)$ is the set

of animals that are observed during the whole interval or part of the interval $(\tau_{m-1}, \tau_m]$. The upper limit of the integral sign $(\tau_m \wedge y_i)$ means minimum of τ_m and y_i.

2.2.3. Prior specification

A priori β_b, $b = 1, ..., P$, λ_{0m}, $m = 1, ..., M$, σ_a^2 and σ_e^2 are assumed to be mutually independent. The prior of β_b was assumed to be improper uniform, *i.e.* $p(\beta_b) \propto 1$ for $-\infty < \beta_b < \infty$ for $b = 1, ..., P$. The priors of the variance components σ_a^2 and σ_e^2 were assumed to be improper uniform, *i.e.* $p(\sigma_a^2) \propto 1$ and $p(\sigma_e^2) \propto 1$ for $0 < \sigma_a^2, \sigma_e^2 < \infty$. *A priori* it was assumed that random variables associated with the piecewise constant steps of the baseline hazard, λ_{0m}, were independent with $p(\lambda_{0m}) \propto \lambda_{0m}^{-1}$, $0 < \lambda_{0m} < \infty$ for $m = 1, ..., M$.

2.2.4. Fully conditional posterior distributions and implementation of the Gibbs sampler

A Bayesian analysis using Gibbs sampling requires a set of starting values of augmented parameters, $\boldsymbol{\psi}$, and sampling from the set of fully conditional posterior distributions of all the elements of this augmented parameter vector $\boldsymbol{\psi}$. We use single updating for all the elements of $\boldsymbol{\psi}$. The fully conditional posterior distribution of every parameter is proportional to the posterior distribution of $\boldsymbol{\psi}$, which up to proportionality is given by the product of the conditional likelihood (2) and the prior of $\boldsymbol{\psi}$. The fully conditional posterior distributions of all the elements of $\boldsymbol{\psi}$ are specified in the following, where for a particular element, φ, of $\boldsymbol{\psi}$ we let $\boldsymbol{\psi}_{\backslash\varphi}$ denote all elements in $\boldsymbol{\psi}$ except φ. The starting values for all the coordinates of $\boldsymbol{\beta}$, $\mathbf{a}$ and $\mathbf{e}$ were set equal to zero and the initial values for the additive genetic and the residual variances were 0.2 and 0.3.

The fully conditional posterior distribution of λ_{0m}, $m = 1, ..., M$, is up to proportionality given by

$$p(\lambda_{0m} \mid \mathbf{y}, \boldsymbol{\delta}, \boldsymbol{\psi}_{\backslash\lambda_{0m}}) \propto \lambda_{0m}^{d(\tau_m)-1} \exp\{-\lambda_{0m}\Omega_{\lambda_{0m}}\},$$

where $\Omega_{\lambda_{0m}} = \sum\limits_{i \in R(\tau_m)} \exp\{a_i + e_i\} \int_{\tau_{m-1}}^{\tau_m \wedge y_i} \exp\{\mathbf{x}_i'(s)\boldsymbol{\beta}\} \, \mathrm{d}s$. This can be recognised as a Gamma distribution with shape parameter $d(\tau_m)$ and scale parameter $\Omega_{\lambda_{0m}}^{-1}$. However the value of λ_{0m} is expected to be close to zero, the lower bound of the domain, and to avoid numerical instabilities, first γ_{0m} was sampled from a log Gamma distribution with parameters $d(\tau_m)$ and $\Omega_{\lambda_{0m}}^{-1}$ (using ARS), next we

let $\lambda_{0m} = \exp\{\gamma_{0m}\}$, this is a realised value from the fully conditional posterior distribution of λ_{0m}. For an animal i with a record, the additive genetic value, a_i, was sampled using ARS from the fully conditional posterior distribution given by

$$p(a_i \mid \mathbf{y}, \boldsymbol{\delta}, \boldsymbol{\psi}_{\backslash a_i}) \propto \exp\left\{a_i\left(\delta_i - \frac{1}{\sigma_a^2}\sum_{\substack{j=1\\j\neq i}}^{N} A^{ij}a_j\right)\right\}$$

$$\times \exp\left\{-\frac{A^{ii}}{2\sigma_a^2}a_i^2\right\}\exp\left\{-\exp\{a_i\}\Omega_i^a\right\},$$

where A^{ij} is the (i, j)th element of the inverse of the additive genetic relationship matrix and $\Omega_i^a = \exp\{e_i\}\int_0^{y_i}\lambda_0(s)\exp\{\mathbf{x}_i'(s)\boldsymbol{\beta}\}\,\mathrm{d}s$. For an animal i without a record, the additive value was sampled from a normal distribution with mean $-\frac{1}{A^{ii}}\sum_{\substack{j=1\\j\neq i}}^{N} A^{ij}a_j$ and variance $\frac{\sigma_a^2}{A^{ii}}$.

The additive variance was sampled from the Inverted Gamma distribution with parameters $\frac{N}{2}-1$ and $\left(\frac{\mathbf{a}'\mathbf{A}^{-1}\mathbf{a}}{2}\right)^{-1}$, $i.e.$ the fully conditional posterior density of σ_a^2 is up to proportionality given by $\left(\sigma_a^2\right)^{-\frac{N}{2}}\exp\left\{-\frac{1}{\sigma_a^2}\left(\frac{\mathbf{a}'\mathbf{A}^{-1}\mathbf{a}}{2}\right)\right\}$.

The bth regression parameter, β_b; $b = 1, ..., P$, was sampled, using ARS, from the fully conditional posterior distribution of β_b, with density up to proportionality given by

$$p(\beta_b \mid \mathbf{y}, \boldsymbol{\delta}, \boldsymbol{\psi}_{\backslash \beta_b}) \propto \exp\{\beta_b d(\beta_b)\}\exp\left\{-\exp\{\beta_b\}\Omega_b^\beta\right\},$$

where $d(\beta_b)$ is the number of animals that died affected by level b of the covariates ($i.e.$ the number of animals that died with $x_{ib}(y_i) = 1$) and

$$\Omega_b^\beta = \sum_{i=1}^{n}\exp\{a_i + e_i\}\int_0^{y_i}\lambda_0(s)I\{x_{ib}(s) = 1\}\exp\left\{\sum_{\substack{p=1\\p\neq b}}^{P} x_{ip}(s)\beta_p\right\}\,\mathrm{d}s. \text{ In } \Omega_b^\beta,$$

$I\{x_{ib}(s) = 1\}$ is the indicator function which takes the value 1 if the bth coordinate of the design vector of animal i at time s, $x_{ib}(s)$, is equal to 1 and 0 otherwise.

The residual value e_i, for $i = 1, ..., n$, was sampled, using ARS, from the fully conditional posterior distribution given by

$$p(e_i \mid \mathbf{y}, \boldsymbol{\delta}, \boldsymbol{\psi}_{\backslash e_i}) \propto \exp\{e_i\delta_i\} \exp\left\{-\frac{e_i^2}{2\sigma_e^2}\right\} \exp\left\{-\exp\{e_i\}\,\Omega_i^e\right\},$$

where $\Omega_i^e = \exp\{a_i\} \int_0^{y_i} \lambda_0(s) \exp\{\mathbf{x}_i'(s)\boldsymbol{\beta}\}\,\mathrm{d}s$.

The residual variance was sampled from the inverted Gamma distribution with parameters $\frac{n}{2} - 1$ and $\left(\frac{\mathbf{e}'\mathbf{e}}{2}\right)^{-1}$, *i.e.* the fully conditional posterior density of σ_e^2 is up to proportionality given by $\left(\sigma_e^2\right)^{-\frac{n}{2}} \exp\left\{-\frac{1}{\sigma_e^2}\left(\frac{\mathbf{e}'\mathbf{e}}{2}\right)\right\}$.

In the model with time dependent covariates, all of the fully conditional posterior distributions of elements of $\boldsymbol{\psi}$ were very similar to those obtained by Korsgaard *et al.* [13] in the model without time dependent covariates. The main difference is that the integral terms in the $\Omega's$ become more involved and computationally more time consuming in models with time dependent covariates.

This Gibbs sampler was implemented in a Fortran 90 programme, in which we used the subroutines by Gilks and Wild [1] for adaptive rejection sampling.

After 100 000 rounds of the Gibbs sampler, 9000 samples of model parameters were saved with a sampling interval of 100; so the total length of the chain was 1 000 000 rounds.

3. RESULTS

3.1. Variances components

Table II shows marginal posterior summary statistics for the variance components and for the heritabilities on the log frailty scale ($h_{nor}^2 = \sigma_a^2/(\sigma_a^2 + \sigma_e^2)$) and on the $\log(g_i(t))$ scale ($h_{\log}^2 = \sigma_a^2/(\sigma_a^2 + \sigma_e^2 + \frac{\pi^2}{6})$) where $g_i(t)$ is given by $\int_0^t \lambda_0(s) \exp\{\mathbf{x}_i'(s)\boldsymbol{\beta}\}\mathrm{d}s$ [14]. Note that, according to [14] $\log(g_i(T_i)) = -a_i - e_i + \varepsilon_i$, where ε_i follows an extreme value distribution with $\mathrm{Var}(\varepsilon_i) = \pi^2/6$, and with ε_i being independent of $\mathbf{a}$ and $\mathbf{e}$.

For the heritability on the $\log(g_i(t))$ scale the probability of a value greater than 0.01 was 99.34%.

The variance components and the two different heritabilities (variance ratios) show good mixing properties according to the effective sample sizes obtained for the chains for these parameters. Because h_{nor}^2 and $h_{\log}^2$ are functions of the variance components, these parameters are expected to show poor

Table II. Posterior mean (PM), posterior standard deviation (PSD), smallest interval with 95% of the density (95% highest posterior density region) (95–SI), effective sample size (ESS) and lag 1 and lag 10 autocorrelations for the chains of the variance components σ_a^2 and σ_e^2, and the heritabilities h_{nor}^2 and $h_{\log}^2$.

	PM	PSD	95–SI	ESS	Lag 1	Lag 10
σ_a^2	0.247	0.115	0.046–0.482	717.8	0.784	0.203
σ_e^2	0.690	0.372	0.003–1.363	925.9	0.734	0.169
h_{nor}^2	0.300	0.163	0.017–0.623	464.8	0.847	0.319
$h_{\log}^2$	0.095	0.040	0.018–0.174	787.3	0.783	0.205
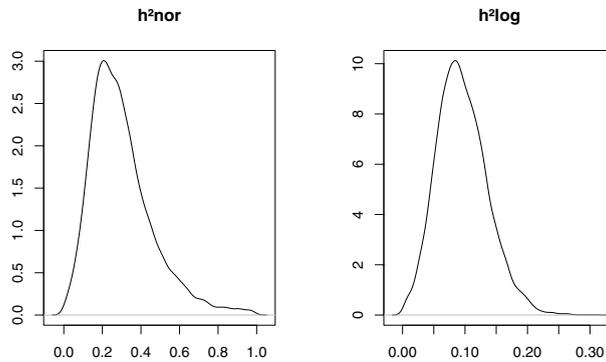
Figure 1. Marginal posterior distributions for h_{nor}^2 and $h_{\log}^2$.

mixing properties. For these parameters convergence was assessed by visual inspection of the trace plot of the chains (not shown) and by the Geweke's test [11]. The Z-score from this test was -1.18 for h_{nor}^2 and 1.23 for $h_{\log}^2$. These values indicate that the test was passed satisfactorily, since the Z-scores are within the interval $(-1.96, 1.96)$.

The marginal posterior distributions of h_{nor}^2 and $h_{\log}^2$ are both right skewed implying that the mode is lower than the mean (Fig. 1).

3.2. Regression parameters

Based on the sampled values of β, estimable functions (γ) for the different levels of all the factors were computed in a similar way to how the least squares means are computed in linear models. After the computation of these estimable functions the risk for each one of these γ_p was calculated as $\exp\{\gamma_p\}$. Note that the ratio between the risks of two different levels will be the relative risk between these two particular levels.

Table III. Posterior means (posterior standard deviations) of the risks associated with the different levels of OPPT*PS[1].

	Pregnant	Lactating	Empty	Pr.&La.
OPPT-1	0.463 (0.094)	0.061 (0.039)	0.591 (0.308)	- -
OPPT-2	0.132 (0.074)	0.038 (0.020)	0.326 (0.127)	0.018 (0.013)
OPPT-3	0.033 (0.017)	0.018 (0.009)	0.315 (0.148)	0.002 (0.001)

[1]OPPT: order of the positive pregnancy test; PS: physiological state.
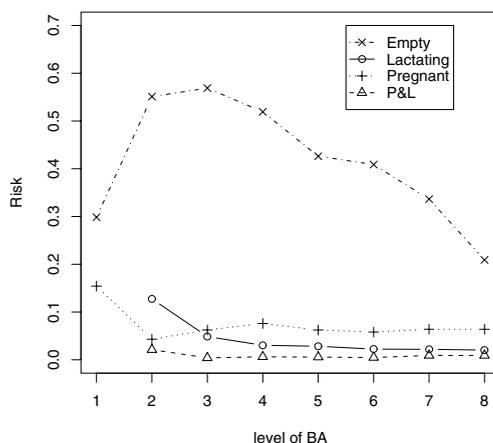
Figure 2. Posterior means of risks associated with the different levels of PS*BA.

The posterior mean and the posterior standard deviation for the risks associated with the different levels of OPPT*PS are given in Table III. For a given level (Pregnant, Lactating, Empty or Pregnant&Lactating) of physiological state, the animals with 1st order of positive pregnancy test (OPPT-1) always had the highest risk followed by does with 2nd (OPPT-2) and higher (OPPT-3) order of positive pregnancy test. Furthermore Table III shows that for a given level (1st, 2nd or ≥3 order of positive pregnancy test) of OPPT, the physiological state Empty was always the level with the highest relative risk followed by Pregnant, Lactating and finally Pregnant&Lactating. However because of the interaction between OPPT and PS, the magnitude of the differences between levels of PS is different for different levels of OPPT and similarly the magnitude of the differences between levels of OPPT is different for different levels of PS.

In Figure 2 the posterior means of the risks associated with the different levels of the time dependent interaction between physiological state and number born alive, PS*BA, are plotted for each of the four levels of physiological

Table IV. Posterior standard deviations of the risks associated with the different levels of PS*BA[1].

	Pregnant	Lactating	Empty	Pr.&La.
BA=0	0.095	- -	0.143	- -
BA=1–2	0.035	0.067	0.136	0.020
BA=3–4	0.037	0.027	0.319	0.005
BA=5–6	0.041	0.016	0.261	0.005
BA=7–8	0.033	0.015	0.209	0.004
BA=9–10	0.030	0.011	0.195	0.003
BA=11–12	0.033	0.011	0.164	0.006
BA= ≥13	0.033	0.011	0.108	0.006

[1]PS: physiological state; BA: number born alive.

state, PS. In general this figure shows the same pattern as Table III for PS, in the sense that for a given level of BA the highest risk is obtained for Empty followed by Pregnant, Lactating and Pregnant&Lactating. Only for level 2 of BA (1–2 young born alive) the physiological state with the highest risk is Empty followed by Lactating. This might be a sign of sick does. For these lactating does, after a poor kindling, the risk is dramatically increased due to the great physiological effort involved in lactation. For a given level of physiological state, the following is observed: for PS = Empty the highest risk is obtained for the third level of BA (3–4 young born alive) followed by the second level of BA (1–2 young born alive). For levels 4–8 of BA (5 or more young born alive) a tendency of decreasing risk with increasing number of young born alive was observed. Surprisingly, the risk for level 1 of BA (0 born alive) was relatively low. For PS = Pregnant, the highest risk was observed for level 1 of BA (0 born alive). For levels of BA different from 1, the risks were stable. For PS = Lactating, the risk is decreasing with an increasing number of young born alive. For PS = Pregnant&Lactating, the risk associated with the different levels of BA were very low and stable. In rabbits, the pattern that risk is decreasing with increasing litter size was previously observed in [9] and [17]. Table IV shows the posterior standard deviations for the risk for all the levels of the interaction PS*BA.

3.3. Baseline hazard function

The posterior mean of the cumulative baseline hazard function was computed based on posterior means for each of the steps in the baseline hazard function. If the baseline hazard function is that of a Weibull distribution, then
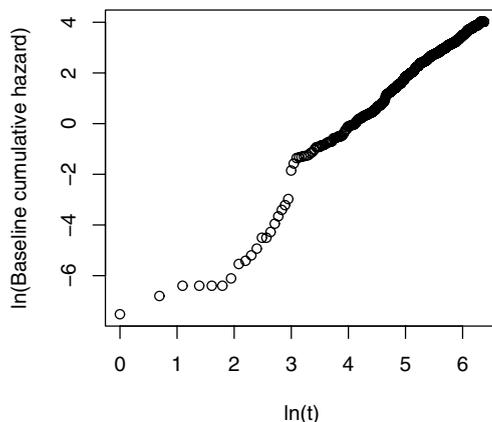
Figure 3. Log posterior mean of $\Lambda_0(t)$ (Baseline cumulative hazard) *versus* ln(t) (days).

the natural logarithm of the cumulative baseline hazard function is a linear function of natural logarithm of time. This is not the case for longevity of our rabbit does, see Figure 3. Two separate regions can be observed, before and after the first parturition, around 19 days after the first positive pregnancy test (ln(19) ≈ 2.94). The same conclusions were obtained by Sánchez *et al.* [17], using the crude Kaplan-Meier estimator for the survival function, although in this previous work the beginning of the productive life was at the first mating.

4. DISCUSSION

This study shows there is additive genetic variation for longevity of rabbit does, and thereby it may be possible to improve longevity by genetic selection. The marginal posterior mean of the additive genetic variance of rabbit doe longevity was 0.247 and the 95% highest posterior density region was 0.046−0.482 (Tab. II). However, despite the additive genetic variation, the marginal posterior mean of $h^2_{\log}$ (this heritability takes into account all the phenotypic variation on the $\log(g_i(t))$ scale) was 0.095, which is a low value. This value could be biased since in the analysis the fact that our population was being selected for litter size at weaning was not taken into account. However the genetic correlation between litter size at weaning and functional longevity has been estimated as very low and not significantly different from zero, thus, this bias is expected to be negligible. In rabbits, low values of heritability of longevity were also obtained by Garreau *et al.* [9], but with a different definition of the trait and another survival model.

The marginal posterior mean of the variance of the residual effect in log-frailty was 0.690 which is almost 3 times higher than the additive genetic variance. This implies that the individual heterogeneity due to environmental unobserved effects and non additive genetic effects is more important compared with the individual heterogeneity due to additive genetic effects.

The data set in this analysis was previously analysed using a different model, a semiparametric log-Normal animal frailty model without residual in log-frailty using the *Survival Kit* [7], and with the covariates given in the present paper. With this model the marginal posterior mean of σ_a^2 was 0.204. This estimate is 17.4% lower than the estimate obtained from the animal model applied in this paper (0.247). Damgaard [4] observed in a simulation study that some bias in the estimation of model parameters exists when there is further individual heterogeneity but it is ignored. Regarding the error in the estimation of the additive genetic variance, it was higher in the model with a residual in log-frailty than in the model without this effect.

The estimates of the effects associated with the different levels of the covariates (β) were larger in absolute value in the animal model with residual term in log-frailty (marginal posterior means) than the ones from the animal model without a residual term in log-frailty (joint posterior mode). The marginal posterior standard deviations for β were larger in the model with the residual term in log-frailty than the standard error reported by the *Survival Kit*.

To our knowledge there is only one study in which the animal model is implemented using the *Survival Kit* to infer the genetic variance of a longevity trait [5]. In order to check the behaviour of the Cox programme of the *Survival Kit* in the estimation of additive genetic variance in animal models, a simple simulation study was carried out. We simulated a population with the same structure as ours, using model (1) but without a residual in log-frailty and fixing the ordered survival times to every two days (to define the steps in the baseline hazard function). The model included a time-independent covariate with two levels and an additive genetic effect with a variance of 0.2. Thirty-five replicates were generated, with a censoring level ranging from 45% to 55%, and the posterior modes, the posterior means and the posterior standard deviation of the genetic variance were estimated. The mean for the posterior means was 0.200 and their standard deviation was 0.050, the mean for the posterior modes was 0.181 and their standard deviation was 0.049, and the mean of the posterior standard deviations was 0.067 and their standard deviation was 0.009. Thus, the posterior mean could be considered an unbiased estimator of the additive genetic variance and the posterior mode as a downward biased estimator of the additive genetic variance in animal models. We have to mention

that in the case of rabbit populations the females are normally closely related and this situation is not the case in other species where the females are not closely related (dairy cattle). Thus, it seems that this particularity of the rabbit populations makes it possible to use the Laplacian integration implemented in the *Survival Kit* to infer the additive genetic variance using animal models.

One drawback with a Bayesian analysis using Gibbs sampling is often the computing time. The application in this study with 2400 records of longevity, corresponding to around 50 000 elementary records and a pedigree file of 3031 animals, took 10.8 days of CPU time in a SGI Altix 3000 system. However we have illustrated that it is possible to carry out a Bayesian analysis using Gibbs sampling in the semiparametric log-Normal animal frailty model with time dependent covariates and a residual effect in log frailty. An advantage of using animal models for survival data is that all additive genetic relationships are taken into account, and biased estimation due to incorrect assumptions concerning the population structure is avoided. With still increasing computer power, animal models for survival data can be expected to be applicable on large data sets within the near future.

ACKNOWLEDGEMENTS

REFERENCES

[1] Adaptive rejection sampling (to generate random variables) from log-concave density functions (42/1993) http://lib.stat.cmu.edu/apstat/287 [consulted: 27 May 2005].

[2] Andersen P.K., Borgan Ø., Gill R.D., Keiding N., Statistical Models Based on Counting Processes, Springer, New York, 1992.

[3] Andersen A.H., Korsgaard I.R., Jensen J., Identifiability of parameters in - and equivalence of animal and sire models for Gaussian and Threshold characters, traits following a Poisson mixed model, and survival traits. Research Report 417 (2000), Department of Theoretical Statistics, University of Aarhus, Denmark.

[4] Damgaard L.H., Quantitative genetic analysis of survival, linear Gaussian and ordered categorical traits, Ph.D. thesis, 2004, The Royal Veterinary and Agricultural University, Frederiksberg C (Denmark).

[5] Diaz C., Chirinos Z., Moreno A., Carabaño M.J., Preliminary analysis of functional longevity in the Avileña Negra Ibérica beef cattle breed, in: Proceeding of the 7th World Congress on Genetics Applied to Livestock Production, 19-23 August 2002, Vol. 29, Montpellier (France) pp. 697–700.

[6] Ducrocq V., Casella G., A Bayesian analysis of mixed survival models, Genet. Sel. Evol. 28 (1996) 505–529.

[7] Ducrocq V., Sölkner J., The *Survival Kit* V3.0, a package for large analyses of survival data, in: Proceedings of the 6th World Congress on Genetics Applied to Livestock Production, 11-16 January 1998, Vol. 27, University of New England, Armidale, pp. 447–450.

[8] Estany J., Baselga M., Blasco A., Camacho J., Mixed model methodology for the estimation of genetic response to selection in litter size of rabbit, Livest. Prod. Sci. 21 (1989) 67–75.

[9] Garreau H., Larzul C., Ducrocq V., Analyse de longévité de la souche de lapins INRA 1077, in: Proceedings of the 9e Journ. Rech. Cunicole, 2001, Paris, France, pp. 217–220.

[10] Geman S., Geman D., Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, IEEE T. Pattern Anal. 6 (1984) 721–741.

[11] Geweke J., Evaluating the accuracy of sampling-based approaches to calculating posterior moments, in: Bernardo J.M., Berger J.O., Dawid A.P., Smith A.F.M. (Eds.), Oxford University Press, Oxford, 1992.

[12] Gilks W.R., Wild P., Adaptative rejection sampling for Gibbs sampling, Appl. Stat. 41 (1992) 337–348.

[13] Korsgaard I.R., Madsen P., Jensen J., Bayesian Inference in the semiparametric log normal frailty model using Gibbs sampling, Genet. Sel. Evol. 30 (1998) 241–256.

[14] Korsgaard I.R., Andersen A.H., Jensen J., Prediction error variance and expected response to selection, when selection is based on the best predictor – for Gaussian and threshold characters, traits following a Poisson mixed model and survival traits, Genet. Sel. Evol. 34 (2002) 307–333.

[15] Ramón J., Rafel O., 1991-1992 - Diez años de gestión global en España, in: Proceeding of Expoaviga, X Jornada Cunícola, 2002, Barcelona, Spain, pp. 113–117.

[16] Rosell J.M., Health status of commercial rabbitries in the Iberian peninsula. A practitioner's study, World Rabbit Sci. 11 (2003) 157–169.

[17] Sánchez J.P., Baselga M., Peiró R., Silvestre M.A., Analysis of factors influencing longevity of rabbit does, Livest. Prod. Sci. 90 (2004) 227–234.

[18] Yazdi M.H., Rydhmer L., Ringmar-Celerberg E., Lundeheim N., Genetic study of longevity in Swedish Landrace sows, Livest. Prod. Sci. 63 (2000) 255–264.