Archivos de Zootecnia
Journal website: https://www.uco.es/ucopress/az/index.php/az/

Genetic evaluation of age at first calving from Brown Swiss cows through survival analysis

Guedes, D.G.P.[1@]; Cunha, E.E.[2] and Lima, G.F.C.[3]

[1]Post-Graduate Program in Animal Production. UFRN/UFERSA. Macaíba. RN. Brazil.
[2]Departament of Cell Biology and Genetics. Federal University of Rio Grande do Norte. Natal. RN. Brazil.
[3]Agricultural Research Company of Rio Grande do Norte (EMPARN). Natal. RN. Brazil.

ADDITIONAL KEYWORDS

Censoring.
Frailty.
Genetic value.
Heifer.
Kaplan-Meier method.

SUMMARY

The age at first calving from Brown Swiss heifers raised in the Semiarid of Brazil was analyzed as well as, the time until the event by the non-parametric method of Kaplan-Meier and the gamma shared frailty model, under the survival analysis methodology. Survival and hazard rate curves associated to this event were estimated and identified the influence of the considered covariates over time. The mean and median ages at the first calving were 987.77 days and 1003 days, respectively, and the significant covariates through the Log-Rank test, in the Kaplan-Meier analysis, were season of birth, calving year, sire (heifer's father) and calving season. In the analysis by frailty model, the genetic values and the frailties of the sires (fathers) for the first calving of their daughters were predicted modeling the hazard function of each heifer as a function of the birth season as a fixed covariate and sire as a random covariate. The frailty followed the gamma distribution. Sires with high and positive genetic values possessed high frailties, which meant a shorter survival time for their daughters at the event, i.e., reduction in their age at first calving.

Avaliação genética da idade ao primeiro parto de vacas Pardo-Suíças por meio da análise de sobrevivência

PALAVRAS-CHAVE ADICIONAIS

Censura.
Fragilidade.
Método de Kaplan-Meier.
Novilha.
Valor genético.

RESUMO

A idade ao primeiro parto de novilhas Pardo-Suíças criadas no Semiárido do Brasil foi analisada como o tempo até o evento, por meio do método não-paramétrico de Kaplan-Meier e do modelo de fragilidade compartilhada gama, sob a metodologia de análise de sobrevivência. Foram estimadas curvas de sobrevivência e de taxa de risco associadas com este evento e identificada a influência das covariáveis consideradas sobre o tempo. As idades média e mediana ao primeiro parto foram 987,77 dias e 1003 dias, respectivamente; as covariáveis significativas pelo teste de Log-Rank, na análise por Kaplan-Meier, foram estação de nascimento, ano de parto, touro (pai da novilha) e estação de parto. Na análise pelo modelo de fragilidade, foram preditos os valores genéticos e as fragilidades dos touros (pais) para o primeiro parto de suas filhas, modelando-se a função de risco de cada novilha em função da covariável fixa estação de nascimento e da covariável aleatório touro. A fragilidade seguiu a distribuição gama. Touros com valores genéticos positivos e altos possuíam fragilidades altas, o que significou um menor tempo de sobrevivência de suas filhas ao evento, ou seja, redução na idade ao primeiro parto delas.

INFORMATION

Cronología del artículo.
Recibido/Received: 29.12.2015
Aceptado/Accepted: 14.11.2016
On-line: 15.04.2017
Correspondencia a los autores/Contact e-mail:
deborahgpguedes@hotmail.com

INTRODUCTION

Some studies have emphasized the importance of reproductive efficiency and its role in the productivity and profitability of dairy cattle herds. Among the traits that are associated with reproductive efficiency, the age at first calving (AFC) is easily measured in the herd and is considered as an indicator of sexual precocity and fertility of heifers. In addition, the AFC is directly related to the animal growth rate, as it is a parameter of expressive economic importance by being related to the beginning of the productive life of the female (Brcko *et al.*, 2010).

In the majority of cattle herds, however, it is common to observe some females without record of the first calving at the moment of evaluation, which can be due to diverse reasons such as disease, physiological immaturity, death or even by the sale of animals, becoming these observations known as censored data (Cunha and Melo, 2012). In genetic evaluations, the methodology of linear mixed models has been traditionally used, which only analyzes the records of females that calved (uncensored data), which cannot follow the normal distribution, discarding the data of animals that have no information.

In this context, the survival analysis has the advantage of taking into account the complete and incomplete information about the individuals that are evaluated as uncensored and censored data, respectively (Allison, 2010) and has been applied to studies on reproductive traits of production animals. In this type of analysis, many different parametric and non-parametric statistical methods can be employed for the treatment of censored data. Among those that have been highly utilized are the non-parametric Kaplan-Meier method or the Product-Limit estimator and the semi-parametric Cox proportional hazards model with shared frailty.

The objective of this study was to evaluate the age at first calving from Brown Swiss heifers, identifying the covariates that influence the time until this event, as well as to predict the genetic value and the frailty of sires for the age at first calving of their daughters, utilizing the survival analysis methodology through the non-parametric Kaplan-Meier estimator and the gamma shared frailty model.

MATERIALS AND METHODS

The data set utilized in this study was collected from a Brown Swiss herd belonging to EMPARN (Agricultural Research Company of Rio Grande do Norte, RN, Brazil), a public company of research in agriculture and livestock production. The herd was raised in the Experimental Station in the Cruzeta city, Rio Grande do Norte, situated in the Central Potiguar mesoregion and the Oriental Seridó microregion. The city is situated at a mean altitude of 26.46 m and at the geographical coordinates of 06°24'43.2" S latitude and 36°47'24.0" W longitude, around 220 km from the state capital.

The climate of the region is characterized as semi-arid, with the wet season lasting until autumn. The mean annual minimum and maximum temperatures, relative humidity and rainfall for the region were estimated in 18.0 °C and 33.0 °C; 64% and 578 mm, respectively (Inmet, 2014). Throughout the months from February to June, the heifers (primiparous cows) were set in a semi-intensive system, fed on native pasture and concentrate formula on-farm; between July and January, an intensive system was adopted using free stall accommodation and a sorghum and elephant grass silage-based diet. 121 heifers were used **(table I)**, daughters of 47 unrelated sires (reproduction by artificial insemination), born in 1993 and between 1995 and 2007, with first calving between 1998 and 2009 and ages at first calving varying from 736 to 2,365 days.

The time (in days) until the first calving (event) occurrence was considered as the dependent variable, expressed on a continuous scale. Under the survival analysis methodology, all the independent (explanatory) variables in the model are known as covariates (Colosimo and Giolo, 2006). Thus, based on studies published in the dairy cattle scientific literature - where the age at first calving was analyzed using linear mixed models -fixed covariates such as the year and season of birth and the year and season of calving were considered, because they could exert some influence on the time variable. The single random covariate was the sire (heifer's father), provided each sire had at least one daughter with its first calving recorded in the herd. All covariates were independent of the time and thus they did not change with time.

The calving seasons were grouped into two categories: season 1 for calving from February to June and season 2 for calving from July to January. Season 1 corresponded to the months with more rainfall in the region, with moderate temperature and humidity levels, while season 2 corresponded to the months of less rainfall with high temperatures and low humidity.

In dairy production systems, the animals that reproduce sooner are the most efficient. In this study, due to the European origin of the Brown Swiss breed, we established that the heifers should have calved by the first time until 36 months of age. Thus, in the survival analysis, the maximum time for recording was fixed at 1,098 days (36 months) since the birth date. Considering that no births occurred prior to 700 days of heifer age, the scale of time was adjusted to discount this long period free of events, taking it as the time of origin. Therefore, when interpreting results, 700 days should be added to heifers' ages.

The *status* variable was created to qualify the time of each heifer as: failure time (uncensored) with status = 1 if she had given birth between 0 and 398 days (700 and 1098 total days) inclusive, and censored time with status = 0 if she had not given birth prior to 398 days (1098 total days). Thus, censored heifers calved at times situated after the censoring-time, resulting in the right-censored data (Colosimo and Giolo, 2006).

The AFC was analyzed in relation to each covariate included in the data set, utilizing the Kaplan-Meier estimator and the gamma shared frailty model, respectively.

The empirical survival function gives us an idea about the distribution of survival times. This function can be obtained using the Kaplan-Meier formula as presented by Allison (2010):

$$\hat{S}_{KM}(t) = \prod_{j:t_j \leq t} \left(\frac{n_j - d_j}{n_j} \right) = \prod_{j:t_j \leq t} \left(1 - \frac{d_j}{n_j} \right)$$

Where $\hat{S}_{KM}(t)$ is the survival function value at time t; t_j represents the ordered failure times since the occurrence of the first failure until the last; n_j represents the number of individuals at risk of the event at t_j ; and $\hat{S}_{KM}(t)$ represents the number of animals that failed at t_j. $\hat{S}_{KM}(t)$ is known as the Kaplan-Meier estimate of survival function or the product-limit estimator.

Table I. **Total number of animals for the covariates sire, year of birth, year of calving, season of birth and season of calving** (Número total de animais para as covariáveis touro, ano de nascimento, ano de parto, estação de nascimento e estação de parto).

Sire				Year of birth		Year of calving		Season of birth		Season of calving	
Sire	Number of daughters	Sire	Number of daughters	Year	Number of animals	Year	Number of animals	Season	Number of animals	Season	Number of animals
1	4	25	1	1993	3	1998	1	1	50	1	49
2	9	26	12	1995	3	1999	14	2	71	2	72
3	1	27	2	1996	4	2000	10				
4	2	28	1	1997	12	2001	8				
5	1	29	2	1998	6	2002	10				
6	5	30	3	1999	16	2003	13				
7	4	31	1	2000	9	2004	17				
8	1	32	1	2001	13	2005	10				
9	1	33	3	2002	16	2006	8				
10	4	34	9	2003	12	2007	14				
11	5	35	2	2004	12	2008	5				
12	1	36	5	2005	5	2009	11				
13	9	37	1	2006	5	-	-				
14	1	38	3	2007	5	-	-				
15	1	39	1								
16	1	40	2								
17	2	41	1								
18	1	42	2								
19	1	43	1								
20	1	44	1								
21	2	45	2								
22	2	46	1								
23	1	47	1								
24	3	-	-								

From $\hat{S}_{KM}(t)$, the empirical hazard function $h(t)$ can be calculated.

The non-parametric Kaplan-Meier estimator shows the effect extension of each covariate on the response variable (time) considered. The non-parametric test utilized to verify the influence of each covariate on the age at first calving of heifers was the Log-Rank test (Mantel, 1966), one of the tests utilized in the survival analysis (Lawless, 1982). Using the forward stepwise sequential method, the chi-square partial statistics were obtained for the Log-Rank test, distributed ordinally from the covariate that showed the highest chi-square (χ^2) value in the general statistics of the univariate analysis.

The PROC LIFETEST from SAS 9.2 (SAS Institute Inc., 2009) was used to obtain the Kaplan-Meier estimator of the survival function $\hat{S}_{KM}(t)$ in order to estimate the survival and hazard rate curves associated with the first calving in this herd, as well as to identify the influence of each covariate on the time until the event.

In time modeling by the gamma shared frailty model, the times of the daughters from the same sire were supposed to possess an association among themselves due to non-observable genetic factors. This justified the inclusion in the model of a random covariate (sire), with a gamma frailty distribution, which was considered in the analysis to take this association into account. Thus, the semi-parametric model of shared frailty was used, which is also considered as an extension of Cox's proportional hazards model.

The final model was chosen after testing the effect of each fixed covariate on the time of calving, as recommended by Collett (2003), in joint analysis with the random effect of sire. In each step of the covariate selection process, the test statistic was obtained using the likelihood ratio test with a chi-square reference distribution with the degrees of freedom (DF) equal to the number of terms excluded (Colosimo and Giolo, 2006).

From this method, the final model was:

$$h_{ijk}(t) = h_0(t)\exp\{estn_i + t_j\} = z_j h_0(t)\exp\{estn_i\}$$

Where $h_{ijk}(t)$ is the hazard function of heifer "*k*", depending on time t until calving, daughter of sire "*j*" and born in season "*i*"; $h_0(t)$ is the unknown baseline hazard function; $estn_i$ is the effect of birth season "*i*" (i = 1 and 2) as the only significant fixed covariate; t_j is the effect of sire (father) "*j*" (j = 1 to 47) as a random covariate, with representing the gamma frailty value (rate of hazard), so that $E(z_j) = 1$ and $Var(z_j) = \xi = \sigma_t^2$.

The heritability (h^2) of AFC was calculated according to the h^2 logarithmic formula as proposed by Ducrocq and Casella (1996):

$$h_{log}^2 = \frac{4\sigma_S^2}{\sigma_S^2 + \frac{\pi^2}{6}}$$

Where $\frac{\pi^2}{6}$ represents the variance of residuals with an extreme value distribution.

Despite Cox's regression model being rather flexible, due to the presence of the non-parametric component in the baseline hazard function, it has not been adjusted for any situation and, as with any other statistical model, requires the use of techniques to evaluate its adequacy (Colosimo and Giolo, 2006). According to these authors, the basic assumption of the model is that the ratio between the hazard rates of two individuals that are at different levels of the same stratum is constant in time and violation of this can lead to serious biases in the estimation of the coefficients of the model.

To verify the assumption of the proportionality of the hazards, the Schoenfeld residuals evaluation was performed, with the residuals defined for uncensored data, i.e. for the failures, and not defined for censored data. To allow for the correlation structure of residuals to be considered, Schoenfeld standardized residuals are frequently used (Colosimo and Giolo, 2006). In this case, the model is expressed by:

$$h(t) = h_0(t)\exp\{x'\beta(t)\}$$

With the restriction that $\beta(t) = \beta$, the hazard proportionality is implied. When $\beta(t)$ is not constant, the impact of one or more covariates on the hazard can vary with time. When the assumption of hazard proportionality is valid, the residual graph against time should be a horizontal line (Colosimo and Giolo, 2006) and should not present sharp tendencies along $g(t) = t$, a function of time. To assist in the detection of a possible error in the assumption of proportional hazards, a curve smoothed with confidence bands was added to the graph.

Because of the subjective conclusions that can be obtained from the interpretation of the graph, the Pearson correlation coefficient (ρ) between the Schoenfeld standardized residuals and $g(t)$ for each covariate was also utilized to verify the assumption of proportional hazards. Colosimo and Giolo (2006) affirmed that ρ values close to zero do not show evidence for the rejection of this assumption.

The analyses were performed using the statistical program R version 3.0.1 (R Development Core Team, 2013), through the *survival* package.

RESULTS

Different quantities of animals were observed for each covariate stratum. Nonetheless, this had no influence on the Kaplan-Meier estimates (Lawless, 1982). For the fixed covariates, the average numbers of heifers and of censored and failed heifers were: 8.64 heifers/ birth year (with 6.14 as failed and 2.5 as censored); 60.5 heifers/ birth season (with 43.0 as failed and 17.5 as censored); 10.08 heifers/ calving year (with 7.17 as failed and 2.91 as censored); and 60.5 heifers/ calving season (43.0 as failed and 17.5 as censored). For the random covariate, the averages were: 2.57 heifers/ sire (1.83 as failed and 0.74 as censored).

For the birth year covariate **(table II),** it was possible to observe different quantities of animals per year: all females born in the years from 1993 and 1995 survived and were censored, while all those born in the years 1996, 2006 and 2007 failed (were uncensored).

Table II. Total number of animals, number of uncensored and censored animals and percentage for the covariate year of birth (Número total de animais, de animais não censurados e censurados e suas respectivas porcentagens para a covariável ano de nascimento).

Year of birth	Number of animals	Number of uncensored animals (failure)	Number of censored animals
1993	3 (2.48%)	0 (0.00%)	3 (100.00%)
1995	3 (2.48%)	0 (0.00%)	3 (100.00%)
1996	4 (3.31%)	4 (100.00%)	0 (0.00%)
1997	12 (9.92%)	10 (83.33%)	2 (16.67%)
1998	6 (4.96%)	5 (83.33%)	1 (16.67%)
1999	16 (13.22%)	11 (68.75%)	5 (31.25%)
2000	9 (7.44%)	5 (55.56%)	4 (44.44%)
2001	13 (10.74%)	10 (76.92%)	3 (23.08%)
2002	16 (13.22%)	15 (93.75%)	1 (6.25%)
2003	12 (9.92%)	4 (33.33%)	8 (66.67%)
2004	12 (9.92%)	9 (75.00%)	3 (25.00%)
2005	5 (4.13%)	3 (60.00%)	2 (40.00%)
2006	5 (4.13%)	5 (100.00%)	0 (0.00%)
2007	5 (4.13%)	5 (100.00%)	0 (0.00%)
Total	121	86	35

Table III. Total number of animals, number of uncensored and censored animals and percentage for the covariate season of birth (Número total de animais, de animais não censurados e censurados e suas respectivas porcentagens para a covariável estação de nascimento).

Season of birth	Number of animals	Number of uncensored animals (failure)	Number of censored animals
1	50 (41.32%)	40 (80.00%)	10 (20.00%)
2	71 (58.68%)	46 (64.79%)	25 (35.21%)
Total	121	86	35

The Log-Rank test was applied to verify the equality between the years of birth, having been detected a significant difference (p<0.0001) by the chi-square statistic (χ^2= 40.9254 with DF = 13).

The season of birth **(Table III)** was also studied as animals born during different seasons of the year can be influenced by different environmental and feeding conditions (Caetano *et al*., 2012). The greater percentage of failures (80.0%) occurred for animals born in season 1, having been recorded the lower number of animals (50) born in this season. Consequently, season 2 had a greater percentage of surviving animals (35.21%) compared to season 1 (20.0%). The Log-Rank test was used to verify the effect of season of birth on the probability of survival of heifers to the first calving, indicating a significant difference (p= 0.0222) between the seasons (χ^2= 5.2291 with DF= 1).

As shown in **Table IV**, none of the heifers calved during 1998, which resulted in 100% survival; the contrary occurred during 2001, when no censoring (survival) was identified. The years 2000, 2002 and 2005, which had 10 heifers each, had failure percentages of 70%, 70% and 90%, respectively. No significant differences were detected (p= 0.6899) between the years of birth (χ^2= 8.2597 with DF= 11), by the Log-Rank test.

A greater number of animals (72) and percentage of failed heifers (73.61%) were observed in the second season of calving **(Table V)**, while in the first season a greater percentage of animals survived (32.65%) if compared to season 2 (26.39%). No significant differences were detected between the two seasons of calving (p= 0.4554), considering the χ^2 statistic of 0.5571 (1 degree of freedom), by the Log-Rank test.

In general, the survival curve of this herd presented a decline from zero to 398 days (700 to 1098 total days), corresponding to the reduction in the probability of survival to the first calving (event) from 1.0 to 0.2893 (100% to 28.93%). In this period, 86 heifers calved by the first time, representing the failure percentage of 71.07%. After that, the survival curve became a straight line parallel to the time-axis, with a fixed probability value of 0.2893 as its last estimate. Overall, 35 heifers were censored and thus survived, corresponding to the survival percentage of 28.93%. Within this percentage, it is possible to observe that four heifers calved after 1000 (1700) days, confirming the number of survivors (four) observed in **Figure 1**.

The mean age at first calving was 287.77 (987.77) days, with the median age of 303 (1003) days. The hazard function, that describes the instantaneous probability with which the event (first calving) occurs, presented increases until 398 (1,098) days **(Figure 2)**, the time in which it reached its maximum value (close to 0.009), but with oscillating behavior. Soon after censoring-time, the hazard decreased rapidly to zero and remained as a straight line parallel to the time-axis, following the pattern of the survival curve.

The influence of each covariate on the time until first calving was evaluated by the forward stepwise sequential method, from the inclusion of the most important covariate (with the highest χ^2 in the univariate analysis) by the Log-Rank test. The year of birth was

Table IV. Total number of animals, number of uncensored and censored animals and percentage for the covariate year of calving (Número total de animais, de animais não censurados e censurados e suas respectivas porcentagens para a covariável ano de parto).

Year of calving	Number of animals	Number of uncensored animals (failure)	Number of censored animals
1998	1 (0.83%)	0 (0.00%)	1 (100.00%)
1999	14 (11.57%)	10 (71.43%)	4 (28.57%)
2000	10 (8.26%)	7 (70.00%)	3 (30.00%)
2001	8 (6.61%)	8 (100.00%)	0 (0.00%)
2002	10 (8.26%)	7 (70.00%)	3 (30.00%)
2003	13 (10.74%)	8 (61.54%)	5 (38.46%)
2004	17 (14.05%)	12 (70.59%)	5 (29.41%)
2005	10 (8.26%)	9 (90.00%)	1 (10.00%)
2006	8 (6.61%)	4 (50.00%)	4 (50.00%)
2007	14 (11.57%)	9 (64.29%)	5 (35.71%)
2008	5 (4.13%)	3 (60.00%)	2 (40.00%)
2009	11 (9.09%)	9 (81.82%)	2 (18.18%)
Total	121	86	35

Table V. Total number of animals, number of uncensored and censored animals and percentage for the covariate season of calving (Número total de animais, de animais não censurados e censurados e suas respectivas porcentagens para a covariável estação de parto).

Season of calving	Number of animals	Number of uncensored animals (failure)	Number of censored animals
1	49 (40.50%)	33 (67.35%)	16 (32.65%)
2	72 (59.50%)	53 (73.61%)	19 (26.39%)
Total	121	86	35

the only covariate that did not influence the time until first calving **(table VI).** The others exercised some degree of influence at levels of 1 or 5% probability.

By the gamma shared frailty model, the time until the first calving was significantly influenced by the season of birth (χ^2 =5.55; DF= 1.00; p= 0.018). Under the proposed model, the regression coefficient for the second season of birth was negative (-0.537), indicating that the risk of occurring the event for heifers from this season was 0.5845 (58.45%) times the risk (from 1.0 or 100%) for heifers from the first season, which was assumed as the reference stratum. Thus, the heifers born in the second birth season had lower probability of calving than heifers from the first season (35.21% and 20.00% of censoring, respectively, as aforementioned). In terms of chance, this meant a reduction of 41.55% in the chance of those heifers to calve relative to the heifers born in the first season.

Besides, for the season of birth, Schoenfeld standardized residuals were obtained over time **(figure 3)** to assess whether the assumption of proportional hazards was obtained over its strata (two seasons of birth), confirming the adopted model. Pearson's correlation coefficient (ρ) was calculated between the Schoenfeld standardized residuals and $g(t)$. The test resulted in ρ = 0.00384 (χ^2 = 0.00137; p= 0.971). According to Colosimo and Giolo (2006), ρ values close to zero validate the assumption of proportional hazards between the strata. Still, because no tendency was evident with time, the assumption of proportional hazards was accepted for the model employed in this study.

The test for frailty (z_j) was not significant (χ^2= 9.67; DF = 7.32; p= 0.23), suggesting that there was no significant association between the times of daughters from the same sire in this herd. However, the frailty variance, assumed as the genetic variance between sires, was not null ($\hat{\sigma}_t^2$= 0.12). With this, it was possible to obtain the heritability value of 0.27, following the formula proposed by Ducrocq and Casella (1996).

The predictions of the genetic values of sires ($\hat{t}_j$), as well as their respective frailty values ($\hat{z}_j$) are reported in **Table VII**. The mean of the genetic values was -0.0097; and the mean of the frailties was 0.9903.

DISCUSSION

The reduction in the age at first calving of heifers is associated with the efficiency and profitability of a dairy cattle production system. Many studies have demonstrated the economical advantage resulting from the increase in the reproductive performance when the heifers begin their reproductive life earlier. Heifers that calved earlier had a greater productive life than later calving heifers; thus, heifers that calve prior to 24 months (732 days) should produce more calves than those that calve at 36 months (1,098 days) of age (Dias *et al*., 2004).

Some traits of economic interest were evaluated in dairy cattle by using parametric models in survival analysis such as the counting of somatic cells to estimate the impact of mastitis (clinical and subclinical) on the functional longevity of American Holstein cows (Caraviello *et al*., 2005); the interval between the final insemination for the genetic evaluation of the conception rate of dairy cows using simulated data (Schneider *et al*., 2005); and the relationship between the reproductive traits and the functional longevity of Holstein, Ayrshire and Jersey cows in Canada (Sewalem *et al*., 2008).

In this study, it was possible to verify through the Kaplan-Meier method that close to zero day (700 days), the origin time, the heifers had greater survival and lower hazard rate to the first calving because they were younger. As time passed, the heifers were aging and started calving, justifying the decrease in the survival and increase in the hazard rate. The survival curve for age at first calving was kept constant after 398 (1,098) days, the censoring-time, as no Kaplan-Meier estimates

Table VI. Inclusion of covariates in order by the forward stepwise method of χ^2 statistics for the Log-Rank test (Sequência de inclusão das covariáveis pelo método *forward stepwise* das estatísticas χ^2 para o teste de Log-Rank).

Covariates	DF	χ^2 joint	Pr> χ^2 joint
Season of birth	1	5.2040	0.0225*
Year of birth	2	6.8407	0.2008^{NS}
Year of calving	3	56.8498	<0.0001**
Sire	4	63.6932	0.0089**
Season of calving	5	69.0183	0.0210*

**p<0.01; *p<0.05.

Table VII. Predicted genetic values $(\hat{t}_j)$ and frailty values $(\hat{z}_j)$ for sire j (1 to 47), considering the gamma shared frailty model (Valores genéticos preditos $(\hat{t}_j)$ e valores da fragilidade $(\hat{z}_j)$ para o touro j (1 a 47) considerando o modelo de fragilidade compartilhada gama).

Sire	$\hat{t}_j$	$\hat{z}_j$	Sire	$\hat{t}_j$	$\hat{z}_j$
1	-0.4786	0.6196	25	0.0484	10.496
2	0.3601	1.4334	26	-0.1664	0.8467
3	0.0852	1.0889	27	-0.0775	0.9254
4	0.1940	1.2141	28	0.0016	10.016
5	-0.1225	0.8847	29	-0.0002	0.9998
6	0.0873	1.0912	30	0.1841	12.021
7	-0.0318	0.9687	31	0.0435	10.444
8	0.1116	1.1181	32	0.0364	10.371
9	0.0246	1.0249	33	-0.2331	0.7920
10	0.0674	1.0697	34	-0.1904	0.8266
11	-0.0717	0.9308	35	0.1502	11.620
12	0.0422	1.0431	36	-0.0332	0.9674
13	-0.0519	0.9494	37	-0.1225	0.8847
14	0.0563	1.0579	38	-0.0238	0.9765
15	0.1124	1.1190	39	-0.1225	0.8847
16	-0.2013	0.8177	40	-0.0466	0.9545
17	-0.1280	0.8798	41	0.0079	10.079
18	0.0451	1.0462	42	0.1533	11.657
19	0.1064	1.1123	43	0.0710	10.736
20	0.0504	1.0517	44	0.0981	11.030
21	0.0438	1.0448	45	-0.1129	0.8932
22	-0.2316	0.7933	46	0.0966	11.014
23	-0.1225	0.8847	47	-0.2013	0.8177
24	0.0375	1.0382			

were reported for censored records using the PROC LIFETEST program from SAS (Allison, 2010).

The heifers born in the second season, when less rainfall occurred, had a greater survival percentage, meaning that they were not the most precocious in the herd. This can be related to a possible feeding deficiency of the mothers of these heifers, as a consequence of a lower availability of pasture at this time of the year. Despite concentrate supplementation, the heifers born in the second season required a greater amount of feed to generate sufficient energy for maintenance and milk yield to feed their calves. Heifers born in the first season were most likely privileged due to the better quality of food provided for their mothers, which at the same time guaranteed more feed to their calves before weaning.

The percentage of censored heifers (28.93%), i.e., those that did not reach their first calving until the established censoring-time, was relatively high. This can be a reflection of the different environmental conditions related to the effects of the statistically significant covariates, season of birth and year and season of calving, that affected positive or negatively the expression of genetic potential for reproductive precocity of heifers in this herd. It should also be considered that the AFC is a quantitative trait, which presented a heritability (h^2) estimate of 0.27 in this herd, by Cox's model on a logarithmic scale (Ducrocq and Casella, 1996).

According to Cunha and Melo (2012), aiming the genetic improvement for AFC in herds, the selection of sires whose daughters calved earlier is optimal, as their survival curves exhibit more accentuated decreases over time due to the sires' genetics that would confer greater reproductive precocity to their progenies.

Proportional hazards models can be extended for the inclusion of a random effect as occurs with linear mixed models. According to Ducrocq (1997), survival mixed models are classically known as frailty models. The shared frailty model is formulated by the intro-

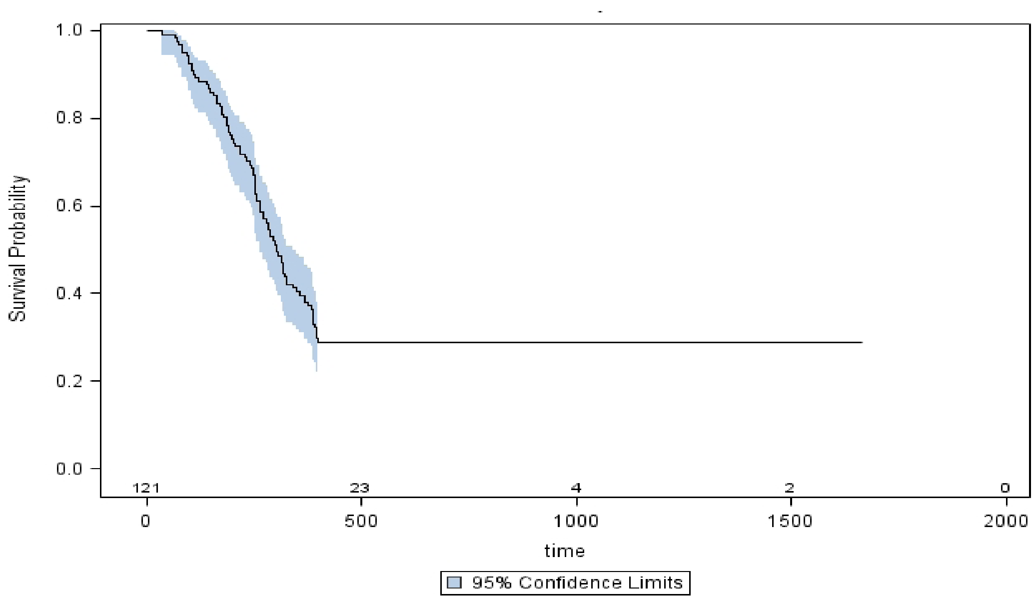

Figure 1. Survival curve, including the number of survivors (inside the x axis) over time (days), with confidence limits of 95% probability for AFC in Brown Swiss cows, from 0 – 398 days (700 to 1098 days) (Curva de sobrevivência, incluindo o número de sobreviventes (parte interna do eixo x) no tempo (dias), com limite de confiança de 95% de probabilidade para a idade ao primeiro parto (IPP) em vacas Pardo-Suíças, de 0 a 398 dias (700 a 1098 dias)).

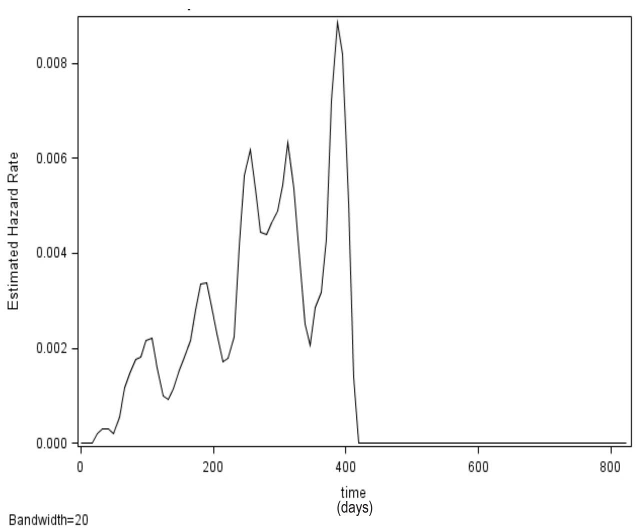

Figure 2. **Epanechnikov kernel-smoothed hazard function curve for the occurrence of first calving in Brown Swiss cows** (Curva da função de risco suavizada-kernel de Epanechnikov para a ocorrência do primeiro parto nas vacas Pardo-Suíças).

duction of a random effect in the Cox model, which acts multiplicatively on the hazard function. Gamma distribution has traditionally been linked to the term of frailty due to its flexibility and mathematical convenience.

In this study, the gamma shared frailty model was utilized to predict the genetic values and the frailties of sires. This model has been used to model multivariate survival data characterized by the existence of natural or artificial groupings to specify the independence between the data observed conditionally on a set of non-observable variables (Colosimo and Giolo, 2006). The frailty represents a random effect that describes the common risk, or the frailty shared by individuals within the same group or family. In general, the idea behind this model is that the sires can present different frailties among themselves, being the frailty of each sire constant over time and common to all its daughters which creates dependence among their event times (Wienke, 2011). Therefore, the daughters of sires with higher frailty values should experience the event of interest (first calving) at earlier times than the daughters of sires with lower frailty values. The surviving heifers were probably daughters of sires that had lower frailty values to the event of interest.

The estimation of genetic values for sires through survival analysis has previously been discussed. Bonetti *et al*. (2009) estimated genetic parameters in a genetic evaluation for the longevity of Brown Swiss sires in Italy using the Weibull proportional hazards model. The predicted genetic values for longevity, expressed as the rate of relative hazard, varied from 0.8 to 1.0. The authors considered the method appropriate for the use and inclusion of the information of sires in genetic improvement programs. Caraviello *et al*. (2004) also estimated the genetic values of Holstein sires for the longevity of their daughters, using the Weibull proportional hazards model. The authors compared the predictions of genetic values obtained by the survival analysis with those obtained by a linear model, which is the most common method used in genetic evalua-

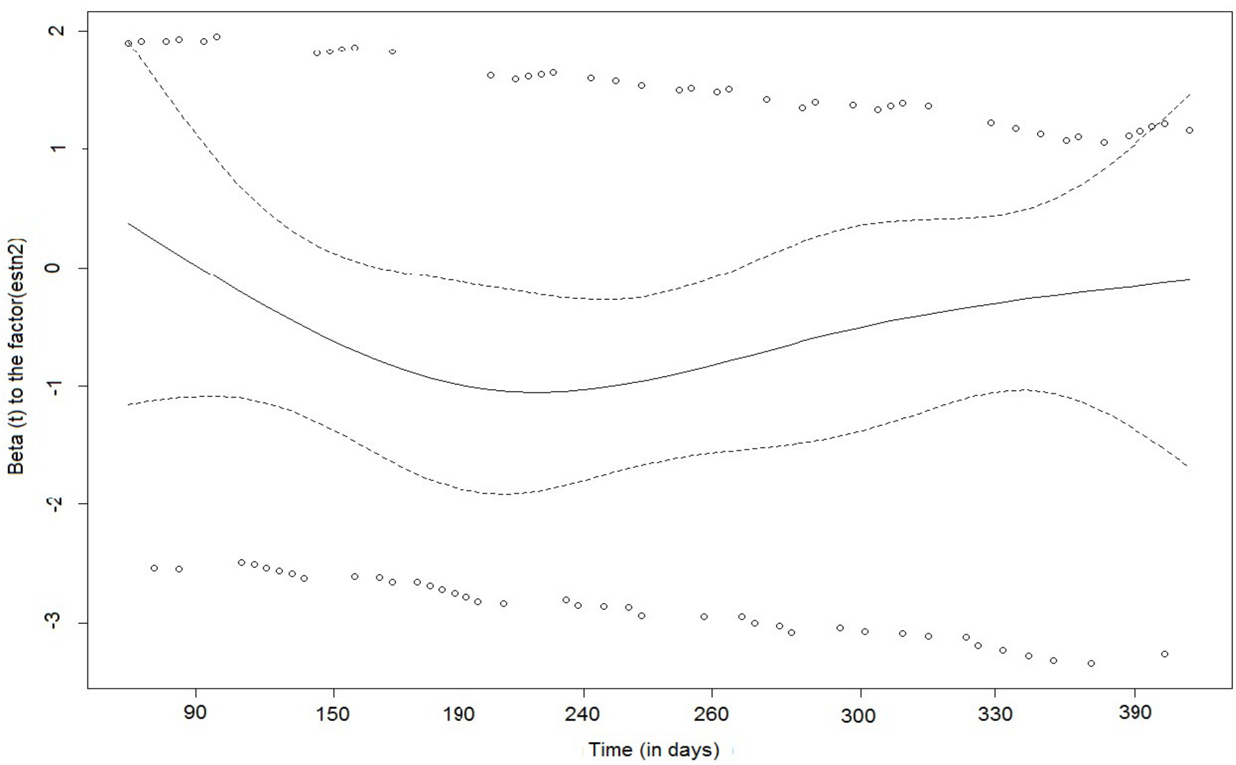

Figure 3. **Schoenfeld standardized residuals for season of birth versus time** (Resíduos padronizados de Schoenfeld para a estação de nascimento versus o tempo).

tion routines, and found that the proportional hazards model resulted in more accurate predictions of the daughters' longevity than the linear model, according to the obtained results .

As frailty acts multiplicatively on the hazard function, sires with a frailty value above 1.0 (reference) had a greater risk (chance) for the event occurrence, which meant that their daughters were more fragile and thus they would have calved earlier **(Table VII)**. Negative genetic values corresponded to frailties below 1.0. Therefore, the sire 1 had the lowest genetic value (-0.4786) associated with the lowest frailty (0.6196), while the sire 2 had the highest genetic value (0.3601) and thus the greatest frailty (1.4334). Therefore, the risk of calving between the daughters of sire 2 was 2.31 times the risk between the daughters of sire 1. As all sires with at least one calved daughter in the herd were evaluated -regardless of whether the daughter record was a failure or censoring – there was a difference in the number of uncensored/censored daughters among the sires, generating an unbalanced data structure. This is allowed in the frailty model in question (Colosimo and Giolo, 2006).

ACKNOWLEDGEMENTS

Thanks to CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) for the fellowship to the first author and to EMPARN by the data set utilized in this study.

REFERENCES

Allison, P.D. 2010. Survival analysis using SAS: a practical guide. 2. ed. SAS Institute Inc. Cary. 324 pp.

Bonetti, O.; Rossoni, A. and Nicoletti, C. 2009. Genetic parameters estimation and genetic evaluation for longevity in Italian Brown Swiss bulls. *Ital J Anim Sci*, 8 (suppl. 2): 30-32.

Brcko, C.C.; Araújo, C.V.; Araújo, S.I.; Rennó, F.P. e Marcondes, C.R. 2010. Estimação de parâmetros genéticos da produção leiteira e idade ao primeiro parto de vacas Pardo-Suíças por meio de inferência bayesiana. *Rev Cien Agra,* 53: 59-53.

Caetano, S.L.; Rosa, G.J.M.; Savegnago, R.P.; Ramos, S.B.; Bezerra, L.A.F.; Lôbo, R.B.; De Paz, C.C.P. and Munari, D.P. 2012. Characterization of the variable cow's age at last calving as a measurement of longevity by using the Kaplan-Meier estimator and the Cox model. *Animal*, 7: 540-546.

Caraviello, D.Z.; Weigel, K.A. and Gianola, D. 2004. Prediction of longevity breeding values for US Holstein sires using survival analysis methodology. *J Dairy Sci*, 87: 3518-3525.

Caraviello, D.Z.; Weigel, K.A.; Shook, G.E. and Ruegg, P.L. 2005. Assessment of the impact of somatic cell count on functional longevity in Holstein and Jersey cattle using survival analysis methodology. *J Dairy Sci*, 88: 804-811.

Collet, D. 2003. Modelling survival data in medical research. 2° ed. Chapman and Hall. London. 395 pp.

Colosimo, E.A. e Giolo, S.R. 2006. Análise de sobrevivência aplicada. Edgard Blücher. São Paulo. 370 pp.

Cunha, E.E. e Melo, T.P. 2012. Análise de sobrevivência não-paramétrica da idade ao primeiro parto em fêmeas Nelore: um estudo de simulação. *Rev Bras Biom*, 30: 305-325.

Dias, L.T.; El Faro, L. e Albuquerque, L.G. 2004. Estimativas de herdabilidade para idade ao primeiro parto de novilhas da raça Nelore. *Rev Bras Zootecn*, 33: 97-102.

Ducrocq, V. 1997. Survival analysis, a statistical tool for longevity data. In: 48th Annual Meeting of the European Association for Animal Production. Vienna. Austria.

Ducrocq, V. and Casella, G. 1996. A Bayesian analysis of mixed survival models. *Genet Sel Evol*, 28: 505-529.

Inmet. 2014. Instituto Nacional de Meteorologia. Banco de dados meteorológicos para ensino e pesquisa. Brazil.

Lawless, J.F. 1982.Statistical models and methods for lifetime data. John Wiley and Sons. New York. 552 pp.

Mantel, N. 1966. Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemother Rep*, 50: 163-170.

R Development Core Team. 2013. R: A language and environment for statistical computing. [3.0.2]. R Foundation for Statistical Computing. Vienna.Austria.

SAS Institute Inc. 2009. SAS/STAT® 9.2 User's guide. 2. ed. SAS Institute Inc. Cary. 7869 pp.

Schneider, M. del P.; Stranberg, E.; Ducrocq, V. and Roth, A. 2005. Survival analysis applied to genetic evaluation for female fertility in dairy cattle. *J Dairy Sci*, 88: 2253-2259.

Sewalen, A.; Miglior, F.; Kistemaker, G.J.; Sullivan, P. and Van Doormaal, B.J. 2008. Relationship between reproduction traits and functional longevity in Canadian dairy cattle. *J Dairy Sci*, 91: 1660-1668.

Wienke, A. 2011. Frailty models in survival analysis. Chapman & Hall/ CRC (Biostatistics Series). Boca Raton. 298 pp.