

Genetic analysis of days from calving to first insemination and days open in Danish Holsteins using different models and censoring scenarios¹

Y. Hou,*† P. Madsen,* R. Labouriau,* Y. Zhang,† M. S. Lund,* and G. Su*²

*Department of Genetics and Biotechnology, Faculty of Agricultural Sciences, Aarhus University, DK-8830, Tjele, Denmark

†College of Animal Science and Technology, China Agricultural University, Beijing, China

ABSTRACT

The objectives of this study were to estimate genetic parameters and evaluate models for genetic evaluation of days from calving to first insemination (ICF) and days open (DO). Data including 509,512 first-parity records of Danish Holstein cows were analyzed using 5 alternative sire models that dealt with censored records in different ways: 1) a conventional linear model (LM) in which a penalty of 21 d was added to censored records; 2) a bivariate threshold-linear model (TLM), which included a threshold model for censoring status (0, 1) of the observations, and a linear model for ICF or DO without any penalty on censored records; 3) a right-censored linear model (CLM); 4) a Weibull proportional hazard model (SMW); and 5) a Cox proportional hazard model (SMC) constructed with piecewise constant baseline hazard function. The variance components for ICF and DO estimated from LM and TLM were similar, whereas CLM gave higher estimates of both additive genetic and residual components. Estimates of heritability from models LM, TLM, and CLM were very similar (0.102 to 0.108 for ICF, and 0.066 to 0.069 for DO). Heritabilities estimated using model SMW were 0.213 for ICF and 0.121 for DO in logarithmic scale. Using SMC, the estimates of heritability, defined as the log-hazard proportional factor for ICF and DO, were 0.013 and 0.009, respectively. Correlations between predicted transmitting ability from different models for sires with records from at least 20 daughters were far from unity, indicating that different models could lead to different rankings. The largest reranking was found between SMW and SMC, whereas negligible reranking was found among LM, TLM, and CLM. The 5 models were evaluated by comparing correlations between predicted transmitting ability from different data sets (the whole data set and 2 subsets, each containing half of the whole data set), for sires with records from at

least 20 daughters, and χ^2 statistics based on predicted and observed daughter frequencies using a cross validation. The model comparisons showed that SMC had the best performance in predicting breeding values of the 2 traits. No significant difference was found among models LM, TLM, and CLM. The SMW model had a relatively poor performance, probably because the data are far from a Weibull distribution. The results from the present study suggest that SMC could be a good alternative for predicting breeding values of ICF and DO in the Danish Holstein population.

Key words: female fertility, genetic evaluation, genetic parameter, model comparison

INTRODUCTION

Reproductive efficiency has a considerable impact on the overall profitability of dairy cattle production. Poor fertility increases costs due to fertility treatments and multiple inseminations, prolongs calving interval, and leads to a high replacement rate due to involuntary culling (Boichard, 1990; Dekkers, 1991; González-Recio et al., 2004). Unfortunately, there is an unfavorable genetic correlation between yield traits and fertility traits (Rauw et al., 1998; Roxström et al., 2001; Van Raden et al., 2004). Consequently, selection for milk production with little or no emphasis on fertility has led to a decline in fertility in modern dairy populations.

Statistical analyses of fertility traits encounter several difficulties mainly because of the presence of censoring and lack of data normality. In practical data collection, if a fertility event for some individuals does not happen by the end of the test period, the records of this fertility event for those cows are usually registered according to the last censoring date (i.e., censored records). Treating censored records as uncensored or excluding them from genetic evaluations will favor sires that have a greater proportion of daughters with censored records, because the censored time is earlier than the time when the fertility event really happens, and the average value of censored records is usually larger than the population mean of the traits. This will lead to biased prediction of breeding values in genetic evaluations.

Received July 16, 2008.

Accepted November 8, 2008.

¹This is a corrected proof. Y. Hou is affiliated with Aarhus University and China Agricultural University.

²Corresponding author: guosheng.su@agrsci.dk

Several methods have been used to deal with censoring in genetic evaluation of fertility traits. One approach is to take a censored record as the true record after adding a penalty (Johnston and Bunter, 1996; Donoghue et al., 2004a, b; Urioste et al., 2007b). Another approach is to generate values for right-censored records from truncated normal distributions (Donoghue et al., 2004a, b; Chang et al., 2006; González-Recio et al., 2006). In addition, frailty survival models have been applied to analyze fertility traits (Schneider et al., 2005, 2006). Another alternative is to use a threshold-linear model, which includes a binary trait indicating censoring status of the observations, and a continuous fertility trait (Donoghue et al., 2004c; Urioste et al., 2007a).

The objectives of this study were to estimate genetic parameters of days from calving to first insemination and days open, using the different models mentioned above, and to evaluate the models regarding stability and predictive ability based on Danish Holstein population data.

MATERIALS AND METHODS

Data

Female fertility data were provided by the Danish Cattle Federation. The data were collected from the Danish Holstein population during the period of insemination from 1992 to 2006. The raw data included fertility records of heifers and cows (the first 3 lactations). Only records from first lactation in the period from 1995 to 2004 were used in the current analysis. The restriction to the period from 1995 to 2004 was imposed to exclude left-censored cows and cows with fertility events in progress (still undergoing inspection of heat and conception). The traits in the analysis were interval between calving and first insemination (ICF) and days open (DO). In this study, DO was defined as interval between calving and conception, conditional on the cow being inseminated. Pregnant status was confirmed by subsequent calving. For cows without subsequent calving, the pregnant status was confirmed by pregnancy test (if available), and was otherwise set to unknown. For ICF, the records of cows that were inseminated were taken as uncensored, and the records of cows that were not inseminated as censored. For DO, the records of cows that conceived were considered as observed, and the records of cows that did not conceive were treated as censored (see more detail about censoring later).

The raw data (including both heifer and cow records) were edited in 2 steps. At the first step, data from cows with age at first calving outside the range of 550 to 1,100 d were excluded. In addition, the data set kept

only records from cows belonging to herds with records in each year and at least 500 records (sum of number of records across heifers and the first 3 parities over the 10 years).

At the second step, the data were further edited for each particular trait in the first lactation. At this step, herd-year classes were required to have a minimum of 5 records. Records were required to be within the normal range, defined as between 20 and 250 d for ICF and between 20 and 365 d for DO. Records with values smaller than the lower limit of the interval were excluded, whereas those larger than the upper limit were replaced with the value of the upper limit and were defined as censored records. Moreover, for the cows (7%) that were not inseminated and were culled after 80 d (about the average of ICF) from calving, ICF records were assumed to be censored at the culling date. The cows (3%) without an insemination but culled before 80 d after calving were deleted, because culling in early lactation was mainly due to disease and low production; that is, nonrandom censoring for fertility traits. Cows with unknown date of confirmed successful insemination were given a censored record of DO, calculated from the date of the last insemination event. For the cows (10%) without an insemination, DO was treated as missing instead of being taken as censored records, because these records were quite different from the censored records of cows that had been inseminated.

After editing, the data contained first-lactation records on 509,512 cows from 12,034 sires and 1,899 herds. The pedigree for sires was built by tracing back as many generations as possible. Thus, the pedigree file included 36,610 animals. The detailed information on real and censored records is shown in Table 1.

Statistical Models

Data of ICF and DO were analyzed using 5 alternative sire models that deal with censored records in different ways.

Linear Gaussian Model In the linear model (LM), the censored records (including both unknown pregnancy status or confirmed not pregnant) were added a penalty of 21 d (the average length of the estrus cycle). For DO, a penalty of 21 d suggested that the cows failing to become pregnant would conceive if given an extra estrus cycle. For ICF, the penalty of 21 d suggested that those cows were expected to be inseminated after 21 d, although the penalty was somewhat arbitrary. The linear Gaussian model to describe ICF and DO was

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zs} + \mathbf{e}, \quad [1]$$

Table 1. Proportion of censored records and means and standard deviations (SD) for days from calving to first insemination (ICF) and days open (DO)

Trait	n	Censored records (%)	Mean ± SD for uncensored records (d)	Mean ± SD for censored records (d)
ICF	509,512	7.4	80.5 ± 40.0	197.9 ± 61.1
DO	475,926	16.6	120.4 ± 66.4	208.2 ± 103.9

where \mathbf{y} is the vector of observations of ICF or DO; \mathbf{b} is the vector of fixed effects including year-month of calving, herd-year of calving, age group (age in months at first insemination before first calving), regression on breed proportion of US Holstein, and regression on total heterozygosity; \mathbf{s} is the vector of the sire genetic effects on daughter performances; \mathbf{e} is the vector of random residuals; and \mathbf{X} and \mathbf{Z} are incidence matrixes associating \mathbf{b} and \mathbf{s} with \mathbf{y} .

It was assumed that $\mathbf{s} \sim N(\mathbf{0}, \mathbf{A}\sigma_s^2)$ and $\mathbf{e} \sim N(\mathbf{0}, \mathbf{I}\sigma_e^2)$, where \mathbf{A} was the sire additive genetic relationship matrix, and σ_s^2 was the sire variance, \mathbf{I} was an identity matrix, and σ_e^2 was residual variance.

Threshold-Linear Model. This approach used censoring status of the observations as a correlated trait to improve the accuracy of genetic parameter estimates and PTA for fertility traits. Thus, the threshold-linear model (TLM) used in the analysis included a threshold model for censoring status and a linear Gaussian model for ICF or DO, where censored records were taken as real records without any penalty. It was assumed that a correlation between the trait of interest and the liability of censoring might alleviate the problem of nonrandom censoring. The bivariate model was

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{l} \end{bmatrix} = \begin{bmatrix} \mathbf{X}_y & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_l \end{bmatrix} \begin{bmatrix} \mathbf{b}_y \\ \mathbf{b}_l \end{bmatrix} + \begin{bmatrix} \mathbf{Z}_y & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_l \end{bmatrix} \begin{bmatrix} \mathbf{s}_y \\ \mathbf{s}_l \end{bmatrix} + \begin{bmatrix} \mathbf{e}_y \\ \mathbf{e}_l \end{bmatrix}, \quad [2]$$

where \mathbf{y} is the vector of observations of ICF or DO, and \mathbf{l} is the vector of liabilities of censoring status (0 for censored records, 1 for uncensored records). The vectors \mathbf{b}_y and \mathbf{b}_l include the same effects as those for LM.

It was assumed that $\begin{bmatrix} \mathbf{s}_y \\ \mathbf{s}_l \end{bmatrix} \sim N(\mathbf{0}, \mathbf{A} \otimes \mathbf{G}_0)$ and $\begin{bmatrix} \mathbf{e}_y \\ \mathbf{e}_l \end{bmatrix} \sim N(\mathbf{0}, \mathbf{I} \otimes \mathbf{R}_0)$, where \mathbf{G}_0 and \mathbf{R}_0 are the covariance matrixes of sire genetic effects and residuals for \mathbf{y} and \mathbf{l} , respectively.

Right-Censored Linear Gaussian Model. The right-censored linear Gaussian model (CLM) was similar to [1], but instead of a penalty of 21 d for censored

records, the model applied data augmentation (Tanner and Wong, 1987) to deal with the right censored records. Thus,

$$\begin{bmatrix} \mathbf{y}_0 \\ \mathbf{w} \end{bmatrix} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{s} + \mathbf{e}, \quad [3]$$

where \mathbf{y}_0 is the vector of uncensored observations of ICF or DO, and \mathbf{w} is the vector of augmented values for censored records with $w_i \geq c_i$ (the censored point). Using Gibbs sampling approach (Sorensen et al., 1998; Guo et al., 2001), w_i was sampled from its fully conditional posterior distribution

$$w_i | \mathbf{b}, \mathbf{s}, \mathbf{w}_{-i}, \sigma_s^2, \sigma_e^2, \mathbf{y}_0 \sim N(\mu_i, \sigma_e^2) I(w_i \geq c_i),$$

where \mathbf{w}_{-i} is the vector of \mathbf{w} without w_i , I is the indicator variable with 1 if $w_i \geq c_i$, or 0 otherwise.

Weibull Proportional Hazard Model. The Weibull proportional hazard model (SMW; Ducrocq and Casella, 1996) used in the analysis was

$$\log[\mathbf{h}(\mathbf{t})] = \log[\lambda_0(\mathbf{t})] + \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{s}, \quad [4]$$

where $\mathbf{h}(\mathbf{t})$ is the vector of hazard functions; $\lambda_0(\mathbf{t})$ is a Weibull baseline hazard function of time \mathbf{t} with a positive scale parameter λ and a shape parameter ρ ; that is, $\lambda_0(\mathbf{t}) = \lambda\rho(\lambda\mathbf{t})^{\rho-1}$; and \mathbf{b} and \mathbf{s} are the same as those in model 1.

Cox Proportional Hazard Model. The fifth model was an approximation to the Cox proportional hazard model (SMC), applying piecewise constant baseline hazard function with constant period length of 21 d. It was assumed that the logarithm of the hazard function took the form

$$\log[\mathbf{h}(\mathbf{t})] = \log[\mathbf{h}_0(\mathbf{t})] + \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{s} + \mathbf{e}, \quad [5]$$

where $\mathbf{h}_0(\mathbf{t})$ is the vector of a piecewise constant function of the time \mathbf{t} ; \mathbf{b} and \mathbf{s} are as defined for model LM; and \mathbf{e} is the vector of the random components

representing the residual and period-specific measurement error. The model was implemented using the representation of proportional piecewise constant hazard models for right-censored data in terms of generalized linear models (Laird and Olivier, 1981). According to this representation, the likelihood function of a piecewise constant hazard model coincides with the likelihood function of a generalized linear model defined with a Poisson distribution, a logarithmic link function, an offset containing the logarithm of the observed time at event or censoring, and a response variable taking value 0 for the censored observations and 1 for the noncensored observations (see also Hougaard, 2001). Here we inserted random components by considering generalized linear mixed models and performing inference by using iterative applications of suitable mixed models (see also Breslow and Clayton, 1993; Wolfinger and O'Connell, 1993; Wolfinger et al., 1994).

The analysis applying CLM with Gibbs sampling algorithm was executed using the program MGP (Korsgaard et al., 2003). The analyses for LM and TLM were carried out using Gibbs sampling procedure, and the analysis for model SMC was conducted using the average information-REML procedure in the DMU package (Madsen et al., 2006; Madsen and Jensen, 2007). Survival Kit V3.12 (Ducrocq and Sölkner, 1998) was used for the analysis based on SMW. In the analysis using Gibbs sampling algorithm, a single chain with a length of 120,000 samples was run. Convergence was monitored by graphical inspection. The first 50,000 samples were discarded as burn-in, and every 10th sample of the remaining 70,000 was saved to estimate the features of the posterior distribution.

Model Comparison

To validate the models, the data set was divided into 2 data sets (A and B) randomly by herds, such that all records from a given herd were in either A or B. Sire transmitting abilities were predicted from the whole data set and the 2 subsets, using the variances estimated from the whole data set. Breed contribution was taken into account when calculating PTA.

Two criteria were used to assess the stability or the prediction ability of the 5 models. The first was Spearman rank correlation between PTA from the whole data set, subset A and subset B, for the sires with records from at least 20 daughters in the whole data. The correlation measures the stability of the models to predict breeding value when the prediction is based on different source of data. In other words, the correlation reflects the degree of reranking in candidates when genetic evaluation is based on PTA from different subsets or different amounts of data information.

The second criterion was the χ^2 statistic for the expected and observed fertility events of the daughters, applying a cross validation. In this cross validation, ICF and DO were divided into 5 intervals with thresholds of 49, 65, 86, and 121 d for ICF, and with thresholds of 68, 97, 134, and 194 d for DO. These thresholds corresponded to the 20th, 40th, 60th, and 80th percentiles of the distribution of the whole data. Then, daughter probabilities for first insemination (for ICF) and for conception (for DO) in each of the 5 periods were estimated using logistic regression of daughter frequency on sire PTA from training data (subset A). The estimated probabilities were used to predict frequency of daughters receiving first insemination or frequency of daughters that conceived in the test data (subset B). Finally, the sum of χ^2 statistic (with correction for continuity) over sires for each stage was calculated as

$$\text{Sum of } \chi^2 = \sum_{i=1}^{ns} \left[\frac{(|O_i - E_i| - 0.5)^2}{E_i} + \frac{(|NO_i - NE_i| - 0.5)^2}{NE_i} \right],$$

where O_i was the observed number and E_i was the expected number of daughters receiving first insemination or daughters that conceived within the specific period, NO_i was the observed number and NE_i was the expected number of daughters receiving first insemination or that conceived out of this period.

Further, the sum of multinomial χ^2 statistic (with correction for continuity) over sires for 5 stages was calculated as

$$\text{Sum of } \chi^2 = \sum_{i=1}^{ns} \sum_{j=1}^5 \left[\frac{(|O_{ij} - E_{ij}| - 0.5)^2}{E_{ij}} \right],$$

where O_{ij} was the observed number and E_{ij} was the expected number of daughters receiving first insemination or daughters that conceived within period j (Sokal and Rohlf, 1995, page 704, formula 17.5C). In the cross validation, the censored records were added to 21 d to calculate expected and observed frequency. The same procedure was applied using subset B as training data and subset A as test data.

RESULTS

The proportions of censored records were 7.4% for ICF and 16.6% for DO (Table 1). Nearly half of the censored records for ICF had values for censoring time larger than the upper limit, whereas about 13% of censored records for DO were larger than the upper limit.

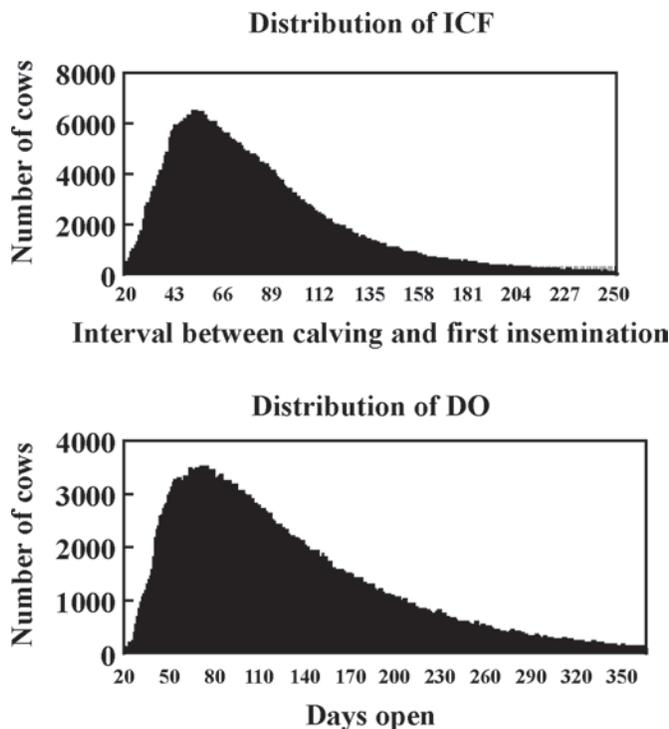


Figure 1. Phenotypic distribution of days from calving to first insemination (ICF) and days open (DO).

These records were replaced with the upper limit value but still treated as censored records. The means of ICF were around 81 d for uncensored records and 198 d for censored records. The means of DO were around 120 d for uncensored records and 208 d for censored records. It was also found that censored records tended to have greater variation than uncensored records.

As shown in Figure 1, both ICF and DO had an asymmetric distribution with a long tail to the right.

For ICF, 80% of the records were in the interval between 20 and 120 d, with the remaining 20% in the interval between 120 and 250 d. For DO, 80% of the records were in the interval between 20 and 194 d, and the others distributed in the interval between 194 and 365 d.

The $\ln[-\ln(\text{Kaplan-Meier estimate})]$ against the natural logarithm of time were plotted (Figure 2). The plot shows a clear deviation from a straight line (a straight line is expected for Weibull distributed data). The coefficient of determination by fitting the $\ln[-\ln(\text{Kaplan-Meier estimate})]$ by regression on the natural logarithm of time was around 0.87 for both ICF and DO. The most serious deviation occurred at the beginnings of the periods in which few cows were inseminated or conceived. These results suggested that a proportional hazard model based on the Weibull distribution fitted the data poorly.

Table 2 shows the estimates of genetic parameters for ICF and DO. The estimates are given as posterior means and standard deviations for models LM, TLM, and CLM, and as modes for SMW and SMC. As expected, CLM gave slightly higher estimates of both sire variance and residual variance than LM and TLM, whereas the estimates from LM and TLM were similar. The estimates of heritability from LM, TLM, and CLM were very similar with a range of 0.102 to 0.108 for ICF and 0.066 to 0.069 for DO. Heritabilities for ICF censoring status (CS_{ICF}) and DO censoring status (CS_{DO}) estimated from TLM were 0.195 and 0.093, respectively. There was a moderate or strong correlation between the fertility trait and its censoring status. The genetic correlation was -0.752 between ICF and CS_{ICF} , and -0.354 between DO and CS_{DO} . The residual correlation was -0.708 between ICF and CS_{ICF} , and -0.411 between DO and CS_{DO} . The estimates of heritability

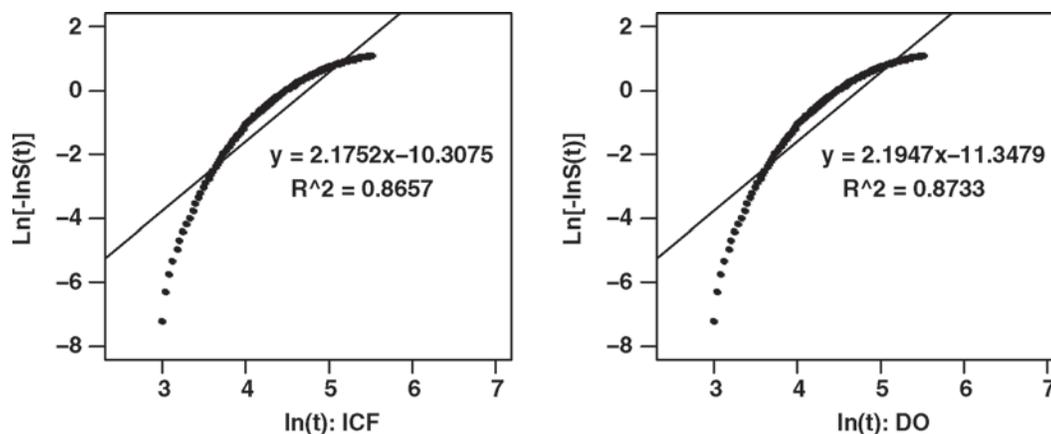


Figure 2. Plot of $\ln[-\ln S(t)]$ against $\ln(t)$. $S(t)$ = Kaplan-Meier estimates of the survivor function at time t , for days from calving to first insemination (ICF) and days open (DO).

Table 2. Posterior means (modes in SMW and SMC) and standard deviations for sire variance (σ_s^2) and heritability (h^2), estimated using linear model (LM), threshold-linear model (TLM), censored linear model (CLM), Weibull proportional hazard model (SMW), and Cox proportional hazard model (SMC)

Trait ¹	Parameter	Model				
		LM	TLM	CLM	SMW ²	SMC ³
ICF	σ_s^2	61.60 ± 3.48	62.20 ± 3.37	70.34 ± 4.04	0.093 ± 0.0038	0.0068 ± 0.0007
	h^2	0.102 ± 0.006	0.108 ± 0.006	0.106 ± 0.006	0.213 ± 0.0082	0.0125 ± 0.0014
DO	σ_s^2	96.52 ± 6.28	95.84 ± 6.29	117.97 ± 7.90	0.051 ± 0.0035	0.0051 ± 0.0006
	h^2	0.066 ± 0.004	0.069 ± 0.004	0.068 ± 0.005	0.121 ± 0.0080	0.0088 ± 0.0011

¹ICF = days from calving to first insemination; DO = days open.

²Heritability for $\log(\text{ICF}_{\text{day}})$ and $\log(\text{DO}_{\text{day}})$ in SMW.

³Heritability for $\log(\text{ICF}_{\text{hazard}})$ of $\log(\text{ICF}_{\text{hazard}})$ in SMC.

were considerably different between the Weibull survival model and the Cox survival model, and between the 2 survival models and the other 3 models, because these models operate in very different scales. Therefore, the heritabilities were not comparable.

Pearson product-moment correlations and Spearman rank correlations between PTA from the 5 alternative models, based on the whole data, were calculated for sires with at least 20 daughters (Table 3). The definitions of PTA differ between the models: PTA in the 2 proportional hazard models (SMW and SMC) reflect the probability, whereas PTA in the other 3 models reflect the days to first insemination (for ICF) or conception (for DO). Therefore, the correlations between PTA from the 2 survival model and from the other 3 models were negative. The patterns of the correlations were the same in both traits. The lowest rank correlation was found between SMC and SMW (0.66 for ICF and 0.76 for DO). The rank correlations were around

−0.83 between SMC and LM, TLM, or CLM for ICF, and around −0.90 for DO. The correlations between SMW and LM, TLM, or CLM were about −0.90 for both ICF and DO. The rank correlations among LM, TLM, and CLM were close to 1 for ICF and greater than 0.97 for DO.

Table 4 displays the Spearman rank correlation between PTA from the whole data set, subset A and B for each of the alternative models, for the sires with at least 20 daughters in the whole data. The highest correlations were found for SMC, the lowest for SMW, and CLM, LM, and TLM were intermediate. For ICF, the correlations between PTA from subset A and B were 0.80 for SMC, 0.27 for SMW, and around 0.50 for LM, TLM, and CLM. For DO, the correlations were 0.78 for SMC, 0.38 for SMW, and around 0.60 for LM, TLM, and CLM. These results indicated that SMC had better stability for breeding value prediction of the 2 traits compared with the other 4 models.

Table 3. Pearson product-moment correlation (above the diagonal) and Spearman rank correlation (below the diagonal) between PTA from different models based on the whole data, for the sires (4,401) with at least 20 daughters (PTA for $\log(\text{ICF}_{\text{hazard}})$ and $\log(\text{DO}_{\text{hazard}})$ in SMW and SMC)

Trait ¹	Model ²	Model				
		LM	TLM	CLM	SMW	SMC
ICF	LM		0.996	0.999	−0.913	−0.840
	TLM	0.996		0.997	−0.920	−0.828
	CLM	0.998	0.997		−0.918	−0.833
	SMW	−0.911	−0.919	−0.917		0.663
	SMC	−0.833	−0.822	−0.828	0.655	
DO	LM		0.998	0.985	−0.914	−0.901
	TLM	0.997		0.973	−0.901	−0.889
	CLM	0.983	0.970		−0.935	−0.909
	SMW	−0.906	−0.891	−0.930		0.772
	SMC	−0.896	−0.884	−0.903	0.759	

¹ICF = days from calving to first insemination; DO = days open.

²LM = linear model; TLM = threshold-linear model; CLM = censored linear model; SMW = Weibull proportional hazard model; and SMC = Cox proportional hazard model.

Table 4. Spearman rank correlation between PTA from the whole data (All), subset A, and subset B, for the sires with at least 20 daughters in the whole data set^{1,2}

Data set	ICF					DO				
	LM	TLM	CLM	SMW	SMC	LM	TLM	CLM	SMW	SMC
All-A	0.836	0.829	0.833	0.763	0.931	0.878	0.880	0.870	0.806	0.931
All-B	0.837	0.828	0.833	0.766	0.936	0.871	0.873	0.864	0.797	0.921
A-B	0.508	0.491	0.497	0.268	0.802	0.620	0.624	0.594	0.384	0.779

¹ICF = days from calving to first insemination; DO = days open.

²LM = linear model; TLM = threshold-linear model; CLM = censored linear model; SMW = Weibull proportional hazard model; and SMC = Cox proportional hazard model.

Table 5 shows χ^2 statistics (calculated from the predicted and observed frequencies in the cross-validation) for each stage and multi-stage, which tested the predictive ability of the models. For ICF, SMC had the smallest multi-stage χ^2 statistic, followed by LM, CLM, and TLM, and SMW had the largest χ^2 statistic. For DO, the order of the 5 models in multi-stage χ^2 statistic was not consistent in the 2 validations. Using PTA from subset A to predict daughter frequency in subset B, the order (from the smallest χ^2 statistic to the largest) was TLM, LM, SMC, CLM, and SMW. Using PTA from subset B to predict daughter frequency in subset A, SMC had the smallest multi-stage χ^2 statistic, SMW had the largest, and LM, TLM, and CLM had similar values. Averaged over the 2 validations, the order (from

the smallest χ^2 statistic to the largest) was SMC, TLM, LM, CLM, and SMW. It was found that the differences in χ^2 statistic between SMC and LM, TLM, or CLM were mainly at the last stage, whereas SMW had a larger χ^2 statistic than the other models at all stages.

DISCUSSION

The present study analyzed ICF and DO in the Danish Holstein population, using 5 models to deal with censored records and lack of normality of the data. The estimates of heritability from LM, TLM, and CLM were somewhat higher than most of estimates in previous reports. The correlations between PTA from different models were far from unity. The stability and predictive

Table 5. Sum of χ^2 statistics over the sires with at least 20 daughters in test data, calculated from the predicted and observed frequency of daughters getting first insemination or conception

Trait ¹	Cross validation	Stage (d)	Model ²				
			LM	TLM	CLM	SMW	SMC
ICF	A predicted B (n = 1,697)	20-49	2,699	2,783	2,731	2,887	2,663
		50-65	1,390	1,399	1,391	1,425	1,372
		66-86	1,346	1,344	1,348	1,376	1,318
		87-121	1,748	1,775	1,758	1,777	1,785
		122-250	2,636	2,718	2,661	3,091	2,512
		Multi-stage	7,836	7,999	7,894	8,440	7,700
	B predicted A (n = 1,770)	20-49	2,948	3,025	2,979	3,249	2,888
		50-65	1,601	1,618	1,602	1,657	1,605
		66-86	1,435	1,438	1,438	1,466	1,401
		87-121	1,734	1,767	1,744	1,839	1,753
		122-250	2,853	2,944	2,879	3,290	2,715
		Multi-stage	8,448	8,627	8,506	9,202	8,284
DO	A predicted B (n = 1,449)	20-68	1,832	1,818	1,905	1,990	1,828
		69-97	1,138	1,136	1,137	1,188	1,146
		98-134	1,119	1,120	1,117	1,137	1,111
		135-194	1,128	1,130	1,133	1,141	1,164
		195-365	1,957	1,937	2,011	2,349	1,981
		Multi-stage	5,742	5,715	5,846	6,265	5,778
	B predicted A (n = 1,484)	20-68	1,943	1,934	1,965	2,064	1,928
		69-97	1,166	1,169	1,165	1,196	1,146
		98-134	1,075	1,079	1,069	1,108	1,061
		135-194	1,223	1,231	1,213	1,240	1,215
		195-365	1,914	1,912	1,900	2,231	1,857
		Multi-stage	5,837	5,836	5,834	6,252	5,757

¹ICF = days from calving to first insemination; DO = days open.

²LM = linear model; TLM = threshold-linear model; CLM = censored linear model; SMW = Weibull proportional hazard model; and SMC = Cox proportional hazard model.

ability of the models were validated by the correlation between PTA from different data sets (the whole data, and subsets A and B), and χ^2 statistic based on predicted and observed daughter frequencies using a cross validation. In general, SMC seemed to perform better than the other 4 models. The LM, TLM, and CLM had similar stability and predictive ability, whereas SMW had a relatively poor performance. It should be noted that no statistical procedure was exerted to test the significance of the difference in performance between these models.

Variance Components and Heritability

The estimates of both additive genetic and residual variances from LM and TLM were the same and were somewhat lower than those obtained from CLM. The difference could be mainly attributed to the augmented records with values greater than the upper limit. In LM and TLM, the records with values larger than the upper limit (250 d for ICF and 365 d for DO) were replaced with the upper limit. But in CLM, those records were taken as censored records with the upper limit as a censored point, and the Markov chain Monte Carlo procedure generated the latent values from right-truncated normal distribution. In addition, a penalty of 21 d was added to the censored records in LM. It is an alternative means, albeit not satisfactory, to account for censoring. A cow may need more than 21 d to receive first insemination or to conceive. This suggests that LM and TLM could slightly underestimate the true variation. However, the estimates of heritability were the same using the 3 models. When analyzing DO in Spanish Holsteins, González-Recio et al. (2006) found that CLM provided greater variances than LM (without penalty), but also provided greater heritability. Urioste et al. (2007a) analyzed days to calving in beef cattle and showed that heritabilities estimated using LM and CLM were similar, whereas lower heritability was estimated using TLM by assigning censored records as missing. In the present study, the Weibull proportional hazard model produced an unconvincingly high heritability estimate, indicating that the model did not fit the data well. Estimates of heritability for log(hazard) from model SMC for ICF and DO were 0.013 and 0.009, respectively. The estimates of heritability from SMC could not be directly compared with those from other models because of different scales.

The estimates of heritability using LM, TLM, and CLM were around 0.106 for ICF and 0.068 for DO, which were greater than most of estimates in previous reports. For ICF, Andersen-Ranberg et al. (2005) reported a heritability of 0.030 in Norwegian dairy cattle with sire LM. For DO, Dechow et al. (2004) showed

a heritability of 0.040 in US Holsteins using sire LM. González-Recio et al. (2006) presented a heritability of 0.053 for DO in Spanish Holsteins using sire LM excluding censored records, 0.056 using sire LM with censored records but without any penalty, and 0.075 using sire CLM.

There may be several possible reasons for various estimates of heritability in different studies. The most important reason could be the criteria on which data editing is based; for example, the definition of the normal range of observations, including or excluding censored records, taking extreme records as missing, or replacing them with the upper (or lower) limit. Oseni et al. (2004) studied the effects of editing criteria on the estimated genetic parameters for DO in US Holsteins, and found that there was an increase in heritability as the upper limit increased from 150 to 250 d in Florida and North Carolina, and little change from 250 to 365 d. E. Norberg (Faculty of Agricultural Sciences, Aarhus University, Tjele, Denmark; personal communication) estimated a heritability of 0.060 for ICF using an animal model, based on data from the same population as the present study, but excluding censored records and defining a normal range of records from 20 to 200 d. Moreover, ICF and DO were largely influenced by management, especially the farmer's decision on the voluntary waiting period. Management in Denmark was relatively similar, which could lead to a reduction of environmental variance. In addition, the various estimates of heritability could reflect a real difference in variation between populations.

In the present study, cows that did not receive first insemination and were culled more than 80 d after calving were given censored records of ICF, and those that were culled after first insemination were given censored records of DO. Therefore, censoring status reflects culling rate to some extent. Bascom and Young (1998) reported that 20% of total culling was due to reproduction, which was taken as the first reason for culling. The present study showed that there was a moderate negative genetic correlation between censoring status (0 = censored, 1 = uncensored) and fertility traits (-0.752 between CS_{ICF} and ICF, -0.354 between CS_{DO} and DO). This indicates that selection for ICF and DO could improve production longevity simultaneously, and vice versa.

Genetic Evaluation Using Different Models

The rank correlations between PTA using LM, TLM, and CLM were all close to 1 for ICF and >0.97 for DO. Urioste et al. (2007b) found a high rank correlation (0.97) for sires with at least 10 daughters between LM and CLM, but a lower correlation between the 2 mod-

els and TLM (around 0.70) for days to calving in beef cattle. This could be because their study considered censored records as missing in TLM. González-Recio et al. (2006) reported a rank correlation of 0.87 between LM without penalty for censored records and CLM for sires that had reliability of PTA higher than 75%. Donoghue et al. (2004a, b) compared LM with penalty for censored records and CLM based on simulated and field data for days to calving in beef cattle and found no difference between the 2 censored data handling techniques in prediction of breeding value.

The rank correlations in PTA between SMW and LM, between SMW and TLM, and between SMW and CLM were very similar (ranged from -0.89 to -0.93). In contrast, González-Recio et al. (2006) reported that the rank correlation between PTA from SMW and PTA from CLM was much higher than that between SMW and LM. So far, we have not found any reports on genetic evaluation of fertility traits using a Cox proportional hazard model as described in the present study.

The 5 models were validated by a comparison on the rank correlation between sire PTA from the whole data, subset A and subset B, and χ^2 statistic from a cross validation. The rank correlations between PTA from different data sets reflect the stability of the models in genetic evaluation. The highest rank correlation was for SMC, followed by LM, TLM, CLM, and SMW, in descending order. This suggests that based on different data sets or different amounts of data information, SMC could result in less reranking of candidates compared with the other 4 models.

The rank correlations between PTA from different data sets were similar for LM, TLM, and CLM, indicating no significant difference between the 3 models regarding stability in genetic evaluation. On the other hand, when analyzing DO, González-Recio et al. (2006) found that SMW and CLM were more stable than LM without penalty for censored records. Urioste et al. (2007b) reported that LM with penalty for censored records and CLM had similar stability for genetic evaluation of days to calving in beef cattle, whereas TLM had better stability when considering censored records as missing.

According to the cross validation, SMC had the best predictive ability, especially for ICF. However, CLM and TLM did not show better predictive ability than LM, and SMW had the poorest predictive ability. An argument is the handling of censored records in the cross validation. The censored records have a contribution to PTA. It is not appropriate to move those records from the cross validation data or treat them as uncensored records. In the present cross validation, they were added a constant penalty of 21 d and then treated as uncensored records. This process of handling

censored records is the same as LM, but different from the other models. Thus, the cross validation is more or less favorable for LM, but unfavorable for the other models. It could be one of the reasons why the more sophisticated models (except for SMC) did not show better predictive ability than LM, based on the present cross-validation procedure.

Many studies have shown that Weibull proportional hazard model could be better than a conventional linear model in genetic evaluation of fertility traits. In a simulation study, Schneider et al. (2005) reported that predicted breeding value of DO using a Weibull survival model had a higher correlation with true breeding value of conception rate than using a conventional linear model. However, the authors did not present the correlation between predicted and true breeding values of DO. González-Recio et al. (2006) analyzed DO in the Spanish Holstein population using different models and found that the Weibull survival model had better predictive ability of daughter fertility at early stages of lactation, based on a cross-validation procedure. In the present study, however, the Weibull survival model showed relatively poor stability and predictive ability. The most likely reason could be that a Weibull distribution did not describe the presented data well. As shown in Figure 2, the curve of $\ln[-\ln(\text{Kaplan-Meier estimate})]$ against $\ln(\text{time})$ showed a clear departure from a straight line.

There are some issues regarding the treatments of censoring used. One is the handling of ICF for the cows without first insemination. In our analyses, those cows were considered censored for ICF at the culling. However, some of those cows were probably not scheduled to be inseminated at all. This means that the censoring was not random, and a large unfair penalty could be placed on those cows. Therefore, if the information is available, it is necessary to distinguish the cows that are not scheduled for insemination and the cows that are scheduled for insemination but do not show detectable estrus, in terms of the values of censored records. In fact, for fertility traits, censoring is typically informative. The models used in this study cannot solve the issue of informative censoring, although the effects of informative censoring are different in the different models used. For instance, informative censoring would imply in a loss of efficiency of the statistical inference in SMC (Andersen et al., 1997), whereas in LM informative censoring clearly implies biased estimation and prediction.

The second issue is the penalty applied to censored records in LM. First, the penalty of 21 d was arbitrary. Moreover, applying a constant penalty to all censored records does not seem to be reasonable. For ICF, the culling date was taken as the censoring date, but cull-

ing can happen at any stage of returning to estrus or at any phase of estrus. Similarly, for DO, different cows at a given censoring date could be at different biological statuses. Therefore, if LM is to be used for genetic evaluation, it is necessary to investigate how large of a penalty should be added to censored records and how to apply distinct penalties to different cows according to their status.

Five alternative models were applied in this study, each of them having advantages and disadvantages. The LM is simple to implement and requires less computational resources, but cannot deal with censored records appropriately. The CLM can handle censored records appropriately, but has a greater computational demand. The TLM, to some extent, accounts for censored records and nonrandom censoring, but has a larger computational load, too. In addition, LM, CLM, and TLM are based on the assumption of normal distribution of the data, which is not satisfied for fertility data. The SMW and SMC models have the advantage of handling censored records and time-dependent effects appropriately. But SMW requires that baseline hazard function has a form of Weibull hazard function, whereas SMC can fit various data-distributions by relaxing assumptions on the form of the baseline hazard function. However, standard SMC based on a classic frailty model is highly computationally demanding, and has difficulty handling large data sets. Moreover, the statistical inference for frailty models constructed with nonparametric baseline hazard functions is complex and might be less efficient (e.g., full-fledged likelihood inference is not available for those models, and inference is typically based on partial likelihood). In contrast, the version of the SMC used in this study is computationally viable and the statistical inference is based on full likelihood. Therefore, it is preferable for the type of applications considered here. The program for the approximate SMC has been integrated into the DMU package, which is able to deal with large data and is available for multi-trait analysis.

CONCLUSIONS

The rank correlations between PTA from SMC, SMW, and the other 3 models (LM, TLM, and CLM) were far from unity, indicating that different models could lead to different rankings of candidates in genetic evaluation. The largest reranking was found between SMW and SMC, with negligible reranking between LM, TLM, and CLM. Results from model validation showed that SMC had the best stability and predictive ability, followed by CLM, LM, and TLM, whereas SMW had a relatively poor performance. These findings suggest that the Cox proportional hazard model could be a

good alternative to genetic evaluation of ICF and DO in the Danish Holstein population.

ACKNOWLEDGMENTS

The authors acknowledge the Danish Cattle Federation for funding the project, Ulrik Sander Nielsen (Danish Agricultural Advisory Service, Denmark) for preparing the data, and Vincent Ducrocq (Station de Génétique Quantitative et Appliquée, Institut National de la Recherche Agronomique, France) for his kind help in applying the proportional hazard model and using Survival Kit.

REFERENCES

- Andersen, P. K., Ø. Borgan, R. D. Gill, and N. Keiding. 1997. *Statistical Models Based on Counting Processes*. Springer-Verlag, New York, NY.
- Andersen-Ranberg, I. M., G. Klemetsdal, B. Heringstad, and T. Steine. 2005. Heritabilities, genetic correlations, and genetic change for female fertility and protein yield in Norwegian dairy cattle. *J. Dairy Sci.* 88:348–355.
- Bascom, S. S., and J. Young. 1998. A summary of the reason why farmers cull cows. *J. Dairy Sci.* 81:2299–2305.
- Boichard, D. 1990. Estimation of the economic value of conception rate in dairy cattle. *Livest. Prod. Sci.* 24:187–204.
- Breslow, N. E., and D. G. Clayton. 1993. Approximate inference in generalized linear mixed models. *J. Am. Stat. Assoc.* 88:9–25.
- Chang, Y. M., I. M. Andersen-Ranberg, B. Heringstad, D. Gianola, and G. Klemetsdal. 2006. Bivariate analysis of number of services to conception and days open in Norwegian Red using a censored threshold-linear model. *J. Dairy Sci.* 89:772–778.
- Dechow, C. D., G. W. Rogers, L. Klei, T. J. Lawlor, and P. M. Van Raden. 2004. Body condition scores and dairy form evaluations as indicators of days open in US Holsteins. *J. Dairy Sci.* 87:3534–3541.
- Dekkers, J. C. M. 1991. Estimation of economic values for dairy cattle breeding goals: Bias due to sub-optimal management policies. *Livest. Prod. Sci.* 29:131–149.
- Donoghue, K. A., R. Rekaya, and J. K. Bertrand. 2004a. Comparison of methods for handling censored records in beef fertility data: Simulation study. *J. Anim. Sci.* 82:351–356.
- Donoghue, K. A., R. Rekaya, and J. K. Bertrand. 2004b. Comparison of methods for handling censored records in beef fertility data: Field data. *J. Anim. Sci.* 82:357–361.
- Donoghue, K. A., R. Rekaya, J. K. Bertrand, and I. Misztal. 2004c. Threshold-linear analysis of measures of fertility in artificial insemination data and days to calving in beef cattle. *J. Anim. Sci.* 82:987–993.
- Ducrocq, V., and G. Casella. 1996. A Bayesian analysis of mixed survival models. *Genet. Sel. Evol.* 28:505–529.
- Ducrocq, V., and J. Sölkner. 1998. The Survival Kit—A Fortran package for the analysis of survival data. *Proc. 6th World Congr. Genet. Appl. Livest. Prod., Armidale, Australia.* 27:447–448.
- González-Recio, O., Y. M. Chang, D. Gianola, and K. A. Weigel. 2006. Comparison of models using different censoring scenarios for days open in Spanish Holstein cows. *Anim. Sci.* 82:233–239.
- González-Recio, O., M. A. Pérez-Cabal, and R. Alenda. 2004. Economic value of female fertility and its relationship with profit in Spanish dairy cattle. *J. Dairy Sci.* 87:3053–3061.
- Guo, S. F., D. Gianola, R. Rekaya, and T. Short. 2001. Bayesian analysis of lifetime performance and prolificacy in Landrace sows using a linear mixed model with censoring. *Livest. Prod. Sci.* 72:243–252.
- Hougaard, P. 2001. *Analysis of Multivariate Survival Data*. Springer, New York, NY.

- Johnston, D. J., and K. L. Bunter. 1996. Days to calving in Angus cattle: Genetic and environmental effects, and covariances with other traits. *Livest. Prod. Sci.* 45:13–22.
- Korsgaard, I. R., M. S. Lund, D. Sorensen, D. Gianola, P. Madsen, and J. Jensen. 2003. Multivariate Bayesian analysis of Gaussian, right censored Gaussian, ordered categorical and binary traits using Gibbs sampling. *Genet. Sel. Evol.* 35:159–183.
- Laird, N., and D. Olivier. 1981. Covariance analysis of censored survival data using log-linear analysis techniques. *J. Am. Statist. Assoc.* 76:231–240.
- Madsen, P., and J. Jensen. 2007. A user's guide to DMU. Version 6, release 4.7. University of Aarhus, Faculty of Agricultural Sciences, Tjele, Denmark.
- Madsen, P., P. Sørensen, G. Su, L. H. Damgaard, H. Thomsen, and R. Labouriau. 2006. DMU—A package for analyzing multivariate mixed models. Book of Abstracts; Proc. 8th World Congr. Genet. Appl. Livest. Prod., Belo Horizonte, Brazil. Commun. no. 07–06.
- Oseni, S., S. Tsuruta, I. Misztal, and R. Rekaya. 2004. Genetic parameters for days open and pregnancy rates in US Holsteins using different editing criteria. *J. Dairy Sci.* 87:4327–4333.
- Rauw, W. M., E. Kanis, E. N. Noordhuizen-Stassen, and F. J. Grommers. 1998. Undesirable side effects of selection for high production efficiency in farm animals: A review. *Livest. Prod. Sci.* 56:15–33.
- Roxström, A., E. Strandberg, B. Berglund, U. Emanuelson, and J. Philipsson. 2001. Genetic and environmental correlations among female fertility traits and milk production in different parities of Swedish Red and White dairy cattle. *Acta Agric. Scand.* 51:7–14.
- Schneider, M. del P., E. Strandberg, V. Ducrocq, and A. Roth. 2005. Survival analysis applied to genetic evaluation for female fertility in dairy cattle. *J. Dairy Sci.* 88:2253–2259.
- Schneider, M. del P., E. Strandberg, V. Ducrocq, and A. Roth. 2006. Short communication: Genetic evaluation of the interval from first to last insemination with survival analysis and linear models. *J. Dairy Sci.* 89:4903–4906.
- Sokal, R. R., and F. J. Rohlf. 1995. *Biometry, the principles and practice of statistics in biological research*. 3rd ed. W. H. Freeman and Company, New York, NY.
- Sorensen, D. A., D. Gianola, and I. R. Korsgaard. 1998. Bayesian mixed-effects model analysis of a censored normal distribution with animal breeding applications. *Acta Agric. Scand. Sect. A* 48:222–229.
- Tanner, M. A., and W. H. Wong. 1987. The calculation of posterior distributions by data augmentation. *J. Am. Stat. Assoc.* 81:82–86.
- Urioste, J. I., I. Misztal, and J. K. Bertrand. 2007a. Fertility traits in spring-calving Aberdeen Angus cattle. 1. Model development and genetic parameters. *J. Anim. Sci.* 85:2854–2860.
- Urioste, J. I., I. Misztal, and J. K. Bertrand. 2007b. Fertility traits in spring-calving Aberdeen Angus cattle. 2. Model comparison. *J. Anim. Sci.* 85:2861–2865.
- Van Raden, P. M., A. H. Sanders, M. E. Tooker, R. H. Miller, H. D. Norman, M. T. Kuhn, and G. R. Wiggans. 2004. Development of a national genetic evaluation for cow fertility. *J. Dairy Sci.* 87:2285–2292.
- Wolfinger, R., and M. O'Connell. 1993. Generalized linear mixed models: A pseudo-likelihood approach. *J. Statist. Comput. Simulation* 4:233–243.
- Wolfinger, R., R. Tobias, and J. Sall. 1994. Computing Gaussian likelihoods and their derivatives for general linear mixed models. *SIAM J. Sci. Comput.* 15:1294–1310.