

# Bivariate Analysis of Number of Services to Conception and Days Open in Norwegian Red Using a Censored Threshold-Linear Model

Y. M. Chang<sup>1</sup>, I. M. Andersen-Ranberg<sup>2</sup>, B. Heringstad<sup>2,3</sup>, D. Gianola<sup>1,3</sup>, and G. Klemetsdal<sup>3</sup>

<sup>1</sup>Department of Dairy Science, University of Wisconsin, Madison 53706, USA

<sup>2</sup>GENO Breeding and A.I. Association, P.O. Box 5003, N-1432 Ås, Norway

<sup>3</sup>Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, P.O. Box 5003, N-1432 Ås, Norway

## 1. Introduction

Genetic evaluation for fertility has been a difficult task in many aspects, mainly because of the incomplete data recording and lack of proper statistical methods to handle skewness and censored records. Interval traits and success/failure traits within some time interval are commonly used for assessing fertility in dairy cattle (Weigel, 2004).

Roxström *et al.* (2001) found high genetic correlation (0.73) between interval from calving to last insemination and number of inseminations per service period. Records for both traits may be censored e.g. cows that are culled before next calving and for which pregnancy status are unknown. Ignoring censoring can lead to incorrect inferences and produce bias on estimates of genetic parameters (Carriquiry *et al.*, 1987). Loss of information due to incomplete records can be minimized if censoring is considered in genetic analysis.

Our objectives were to infer heritabilities for and genetic correlation between number of services to conception (STC) and days open (DO) using a bivariate censored threshold-linear model, and to estimate genetic change for these two traits in Norwegian Red (NRF).

## 2. Materials and Methods

### 2.1. Data

Records of STC and DO from 1,558,705 first-lactation NRF cows with calving and first

insemination data from 1980 to 2004 were included in the analysis. All cows had information on both traits.

Both first and second crop daughters of the 3373 NRF sires that were progeny tested after 1980 were included. If a cow had veterinary treatment of fertility disorders documented, the record was discarded. Age at first calving was between 83 and 174 weeks, and the period from calving to the time data were collected (February 1, 2005) had to be at least 244 d. Double insemination was defined as a new insemination occurring within 0 to 5 d after first insemination. For each cow, all services in first lactation other than a double insemination were counted. STC was in 5 categories, with 5 or more inseminations grouped together as one category. DO was defined as the number of days from calving to last insemination. Cows without a second calving were censored for both traits. Censoring rate was about 19%.

The distribution of cows over STC categories and their mean DO are given in Table 1. About 63% and 25% of the pregnant cows had 1 and 2 inseminations, respectively. DO increased with the number of services to conception, and censored cows had more days open than non-censored cows. Mean DO increased about 4 days over 20 years, while STC had a mean decrease of about 0.1 services (Fig. 2). Raw means were 1.6 STC and 89.7 DO. A total of 3724 males were included in the sire pedigree file.

**Table 1.** Distribution of pregnant and censored cows over number of services to conception (STC) categories, means and standard deviations for days open (in parenthesis).

STC	Pregnant cows		Censored cows	
1	792,738	(68.5±30.8)	156,476	(75.3±37.0)
2	318,669	(104.5±42.5)	85,775	(116.3±47.9)
3	103,384	(135.1±47.1)	40,396	(145.0±50.1)
4	29,646	(163.2±50.4)	15,604	(169.0±50.7)
≥5	9386	(195.2±53.6)	6631	(200.2±53.1)

## 2.2. Statistical model

### 2.2.1. Censored linear model

Survival analysis has been applied in animal breeding because of its ability of handling censored records. An alternative is a mixed effects model analysis of a censored normal distribution (Carriquiry *et al.*, 1987; Sorensen *et al.*, 1998; Guo *et al.*, 2001). Write the linear mixed effects model as:

$$y_i = x'_i \beta + z'_{h,i} \mathbf{h} + z'_{s,i} \mathbf{s} + e_i,$$

where  $y_i$  is the observed non-censored record of cow  $i$ ;  $x_i$ ,  $z_{h,i}$  and  $z_{s,i}$  are incidence vectors related to location parameters of  $\beta$ ,  $\mathbf{h}$  and  $\mathbf{s}$ , and  $e_i$  is the residual. Unobserved responses for censored records can be augmented using a truncated normal process:

$$\tilde{y}_c \sim N(x'_c \beta + z'_{h,c} \mathbf{h} + z'_{s,c} \mathbf{s}, \sigma_e^2) I(y_c, \infty),$$

where  $y_c$  is the observed censoring time, such that the augmented values are larger than the censoring point. This approach does not accommodate time-dependent covariates, but retains the logic of the infinitesimal model of quantitative genetics.

### 2.2.2. Censored threshold model

The threshold model postulates a mixed effect model in the scale of a latent variable, liability ( $\lambda$ ), for each observation (Gianola, 1982; Gianola and Foulley, 1983). The observation takes the value  $j$  only if  $\lambda$  is greater than or equal to  $T_{j-1}$  and smaller than  $T_j$ , where  $T_{j-1}$  and  $T_j$  are some thresholds. This concept

accommodates situations where records are censored at the last observed point. If an observation is censored at the  $j^{\text{th}}$  category, its liability must be larger than  $T_{j-1}$ . The probability that the observation is censored at the  $j^{\text{th}}$  category is:

$$\begin{aligned} \text{Prob}(y_i = j, \text{censored} | \beta, \mathbf{h}, \mathbf{s}, \mathbf{T}) \\ = 1 - \Phi[T_{j-1} - (x'_i \beta + z'_{h,i} \mathbf{h} + z'_{s,i} \mathbf{s})] \end{aligned}$$

The joint probability of  $N$  non-censored and censored data can be given is

$$\begin{aligned} \text{Prob}(\mathbf{y}, \boldsymbol{\delta} | \beta, \mathbf{h}, \mathbf{s}, \mathbf{T}) \\ = \prod_{i=1}^N \{ \Phi[T_{y_i} - (x'_i \beta + z'_{h,i} \mathbf{h} + z'_{s,i} \mathbf{s})] \\ - \Phi[T_{y_i-1} - (x'_i \beta + z'_{h,i} \mathbf{h} + z'_{s,i} \mathbf{s})] \}^{1-\delta_i} \\ \{ 1 - \Phi[T_{y_i-1} - (x'_i \beta + z'_{h,i} \mathbf{h} + z'_{s,i} \mathbf{s})] \}^{\delta_i} \end{aligned}$$

where  $\boldsymbol{\delta}$  is the vector of censoring indicators, where  $\delta_i$  equals to 0 if a record is not censored and 1 otherwise;  $\Phi(\cdot)$  is the standard cumulative normal distribution function, and

$\mathbf{T} = [T_0, T_1, T_2, \dots, T_J]'$  is the vector of unknown thresholds. The thresholds must satisfy the restriction of  $-\infty = T_0 \leq T_1 \leq T_2 \leq \dots \leq T_J = \infty$ . The first threshold  $T_1$  is set to zero, because the parameter cannot be identified in a probit analysis.

### 2.2.3. Bivariate censored threshold-linear model

A Bayesian bivariate model for an ordinal categorical trait and a Gaussian trait (Foulley *et al.*, 1983), with allowance of censored records for the two traits, was fitted:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_h\mathbf{h} + \mathbf{Z}_s\mathbf{s} + \mathbf{e},$$

where  $\mathbf{y} = [\boldsymbol{\lambda}, \mathbf{y}_o, \tilde{\mathbf{y}}_c]'$ . Here  $\boldsymbol{\lambda}$  is the vector of unobserved liabilities to STC, and  $\mathbf{y}_o$  and  $\tilde{\mathbf{y}}_c$  are the vectors of observed non-censored records and augmented censored records for DO, respectively. The vector  $\boldsymbol{\beta}$  included effects of age at first calving (92 levels) and of month-year at first calving (305 levels), specific to each trait. Further,  $\mathbf{h}$  contained herd-5-year effects (110,410 levels),  $\mathbf{s}$  was the vector of sire transmitting abilities (3724 levels), and  $\mathbf{e}$  was the vector of residuals for the two traits.  $\mathbf{X}$ ,  $\mathbf{Z}_h$  and  $\mathbf{Z}_s$  are the incidence matrices.

Residuals for the two traits were assumed correlated within cows and independent between cows:  $\begin{bmatrix} \mathbf{e}_{\text{STC}} \\ \mathbf{e}_{\text{DO}} \end{bmatrix} \sim N(\mathbf{0}, \mathbf{R}_0 \otimes \mathbf{I})$ , where  $\mathbf{R}_0 = \begin{bmatrix} 1 & \sigma_{e12} \\ \sigma_{e12} & \sigma_{e2}^2 \end{bmatrix}$  is the residual (co)variance matrix and  $\mathbf{I}$  is an identity matrix. The residual variance of the liability to STC was set equal to 1,  $\sigma_{e2}^2$  is the residual variance of DO and  $\sigma_{e12}$  is the residual covariance between liability to STC and DO.

Bounded uniform priors were assigned to each of the elements of  $\boldsymbol{\beta}$ . A multivariate normal prior was used for the herd effects,

$p(\mathbf{h} | \mathbf{H}_0) \sim N(\mathbf{0}, \mathbf{H}_0 \otimes \mathbf{I})$ , where  $\mathbf{H}_0 = \begin{bmatrix} \sigma_{h1}^2 & \sigma_{h12} \\ \sigma_{h12} & \sigma_{h2}^2 \end{bmatrix}$  is the 2 x 2 (co)variance matrix between herd effects for the two traits. Effects of different herds were assumed to be independent, a priori. The vector of sire effects was assigned the multivariate normal prior distribution:

$p(\mathbf{s} | \mathbf{A}, \mathbf{G}_0) \sim N(\mathbf{0}, \mathbf{G}_0 \otimes \mathbf{A})$ , where  $\mathbf{G}_0 = \begin{bmatrix} \sigma_{s1}^2 & \sigma_{s12} \\ \sigma_{s12} & \sigma_{s2}^2 \end{bmatrix}$  is the 2 x 2 (co)variance matrix between sire transmitting abilities, and  $\mathbf{A}$  is the additive relationship matrix between male ancestors. Independent inverse Wishart prior distributions were used for the matrices  $\mathbf{H}_0$  and

$\mathbf{G}_0$ , while a scaled inverse chi-square prior distribution was assigned for the residual variance of DO ( $\sigma_{e2}^2$ ). A bounded uniform prior was used for the residual covariance between STC and DO as  $\sigma_{e12} | \sigma_{e2}^2 \sim U(-\sigma_{e2}, \sigma_{e2})$ .

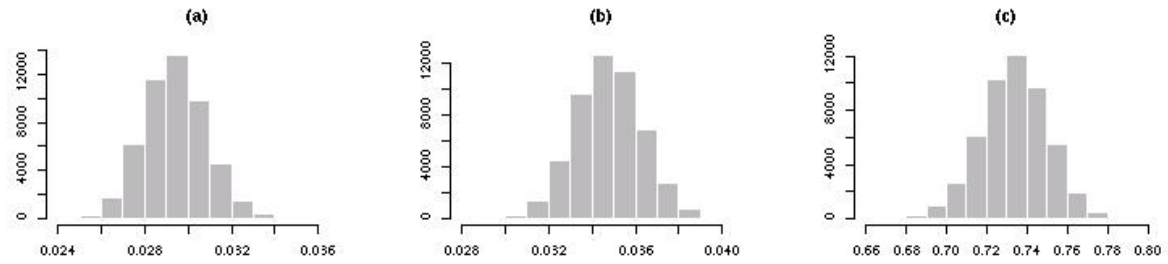
Draws from posterior distributions of the parameters were obtained using a Gibbs sampler, after augmentation of the joint posterior density with the unobserved liabilities to STC and censored DO (Sorensen *et al.*, 1995; Sorensen and Gianola, 2002). The method of Albert and Chib (1997) was used for sampling thresholds, where a logarithmic transformation of the thresholds was adopted as  $\alpha_1 = \log(T_1)$  and  $\alpha_j = \log(T_j - T_{j-1})$ , where  $j=2, \dots, J-1$ . The thresholds can then be obtained as  $T_j = \sum_{i=1}^j \exp(\alpha_i)$ ,  $1 \leq j \leq J-1$ .

Note that  $\alpha_j$  does not have any constraints on order, as in the case of the standard parameterization of thresholds. The first threshold  $T_1$  is fixed at zero, and other thresholds were obtained indirectly by sampling  $\alpha$ s using a Metropolis algorithm. Inferences were based on 50,000 samples, after discarding a burn-in of 10,000 iterations.

### 3. Results and Discussion

Posterior distributions of heritability of liability to STC, of DO, and of the genetic correlation were symmetric and sharp, as shown in Figure 1. Posterior means and standard deviations of the variance components, heritabilities and of correlations between STC and DO are given in Table 2. Low heritabilities, 3%, were found for both traits, which is consistent with previous fertility studies in NRF (Andersen-Ranberg *et al.*, 2005a; 2005b). Wall *et al.* (2003) reported heritability of 0.02 for both 56d non-return and number of insemination per conception, and heritability of 0.04 for days to first service using linear models.

The genetic and residual correlations between STC and DO were 0.73 and 0.69, respectively, which is not surprising since more STC increase DO.



**Figure 1.** Posterior distributions of (a) heritability of liability to number of services to conception (STC), (b) heritability of days open (DO), and (c) genetic correlation between STC and DO.

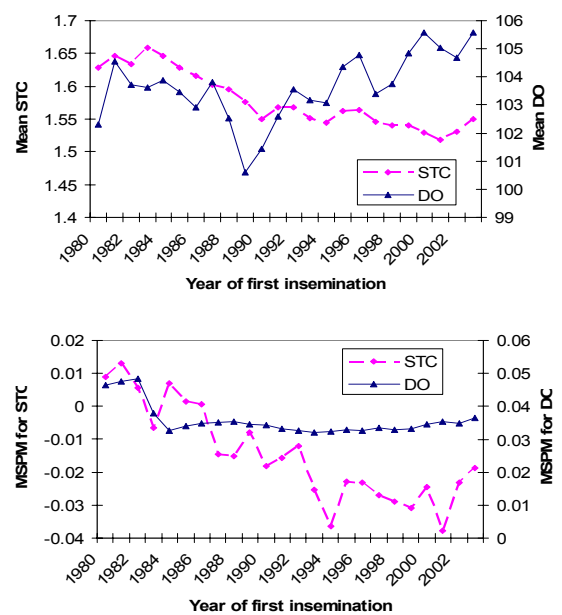
Jamrozik *et al.* (2005) found genetic correlations of 0.92 and 0.96 between number of services and intervals from first service to conception for first and later lactation cows, respectively.

**Table 2.** Posterior means (standard deviations) of variance components, heritabilities, and genetic, herd-5-year and residual correlations between number of services to conception (STC) and days open (DO).

Parameter	STC	DO
Sire variance	0.0074 (0.0004)	21.76 (0.96)
Herd-5-year variance	0.0670 (0.0008)	270.53 (2.30)
Residual variance	1	2480.92 (3.37)
Heritability <sup>1)</sup>	0.03 (0.0014)	0.03 (0.0015)
Genetic correlation	0.73 (0.0164)	
Herd-5-year correlation	0.35 (0.0059)	
Residual correlation	0.69 (0.0005)	

$$^1) h^2 = 4\sigma_s^2 / (\sigma_s^2 + \sigma_e^2)$$

Genetic change for STC and DO in first lactation NRF cows, given as average sire posterior mean by birth-year of daughters is shown in Figure 2. There has been little or no genetic change for DO, while STC shows a decreasing trend. High genetic correlation between non-return rate and number of insemination per conception (-0.94 and -0.88) has been reported (Wall *et al.*, 2003; Jamrozik *et al.*, 2005).



**Figure 2.** Phenotypic and genetic changes of number of services to conception (STC) and days open (DO) in first lactation NRF cows, given as phenotypic means or average sire posterior mean (MSPM) by year of first insemination for daughters.

Since non-return rate has been included in the breeding program for NRF since 1974, the genetic trend for STC is reflecting this favourable correlation between the two traits. Number of services to conception is one of the most important fertility measures but it has not been used so far in the Norwegian breeding program. Methods that take into account censored number of services to conception could increase the accuracy of inferences for this trait, and including STC in a fertility index could improve the genetic evaluation for fertility.

Although the heritability of fertility traits is small, genetic improvement of fertility is still possible if enough weight is put in the total merit index.

## References

- Albert, J.H. & Chib, S. 1997. Bayesian methods for cumulative, sequential and two-step ordinal data regression models. *Technical Report*. Department of Mathematics and Statistics, Bowling Green State University.
- Andersen-Ranberg, I.M., Klemetsdal, G., Heringstad, B. & Steine, T. 2005a. Heritabilities, genetic correlations, and genetic change for female fertility and protein yield in Norwegian Dairy Cattle. *J. Dairy Sci.* 88, 348-355.
- Andersen-Ranberg, I.M., Heringstad, B., Gianola, D., Chang, Y.M. & Klemetsdal, G. 2005b. Comparison Between Bivariate Models for 56-Day Nonreturn and Interval from Calving to First Insemination in Norwegian Red. *J. Dairy Sci.* 88, 2190-2198.
- Carriquiry, A.L., Gianola, D. & Fernando, R.L. 1987. Mixed model analysis of a censored normal distribution with reference to animal breeding. *Biometrics* 43, 929-939.
- Foulley, J.L., Gianola, D. & Thompson, R. 1983. Prediction of genetic merit from data on binary and quantitative variates with an application to calving difficulty, birth weight and pelvic opening. *Genet. Sel. Evol.* 15, 407-424.
- Gianola, D. 1982. Theory and analysis of threshold characters. *J. Anim. Sci.* 54, 1079-1096.
- Gianola, D. & Foulley, J.L. 1983. Sire evaluation for ordered categorical data with a threshold model. *Genet. Sel. Evol.* 15, 201-223.
- Guo, S.F., Gianola, D., Rekaya, R. & Short, T. 2001. Bayesian analysis of lifetime performance in Landrace sows using a linear mixed model with censoring. *Livest. Prod. Sci.* 72, 243-252.
- Jamrozik, J., Fatehi, J., Kistemaker, G.J. & Schaeffer, L.R. 2005. Estimates of Genetic Parameters for Canadian Holstein Female Reproduction Traits. *J. Dairy Sci.* 88, 2199-2208.
- Roxström, A., Strandberg, E., Berglund, B., Emanuelson, U. & Philipsson, J. 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.
- Sorensen, D. & Gianola, D. 2002. *Likelihood, Bayesian and MCMC methods in quantitative genetics*. Springer-Verlag, New York.
- Sorensen, D.A., Andersen, S., Gianola, D. & Korsgaard, I. 1995. Bayesian inference in threshold models using Gibbs sampling. *Genet. Sel. Evol.* 27, 229-249.
- Sorensen, D.A., Gianola, D. & Korsgaard, I.R. 1998. Bayesian mixed effects model analysis of a censored normal distribution with animal breeding applications. *Acta Agric. Scand.* 48, 222-229.
- Wall, E., Brotherstone, S., Woolliams, J.A., Banos, G. & Coffey, M.P. 2003. Genetic evaluation of fertility using direct and correlated traits. *J. Dairy Sci.* 86, 4093-4102.
- Weigel, K. 2004. Improving the reproductive efficiency of dairy cattle through genetic selection. *J. Dairy Sci.* 87, E86-E92.