J. Dairy Sci. 88:3655–3662

Number of Inseminations to Conception in Holstein Cows Using Censored Records and Time-Dependent Covariates

O. González-Recio,[1] Y. M. Chang,[2] D. Gianola,[2] and K. A. Weigel[2]
[1]Departamento de Producción Animal E.T.S.I. Agrónomos – Universidad Politécnica
Ciudad Universitaria s/n 28040 Madrid, Spain
[2]Department of Dairy Science, University of Wisconsin, Madison 53706

ABSTRACT

Three methodologies that accommodate censoring or time-dependent covariates were used to estimate variance components for number of inseminations to conception. Data included 80,071 lactation records and 143,927 artificial inseminations in 47,509 Spanish Holstein cows. Up to 4 inseminations to conception, along with their respective censoring information, were analyzed. An ordinal-censored threshold model (CTM), a sequential threshold model (STM), and a grouped survival analysis via a discrete proportional hazards model (DPH) were implemented. Sire variance estimates on the liability scale were 0.016 and 0.010 for CTM and STM, respectively, and 0.012 for DPH on the logarithmic scale. Heritability estimates on the liability scale were 0.050 and 0.038 with CTM and STM, respectively. All models led to similar rankings of sires, and the strong correlations (0.97 to 0.98) between methodologies suggested robustness in ranking of sires of cows. Service sire variance estimates were 0.021 for both CTM and STM; DPH led to an approximate service sire variance of 0.020. Rankings for service sires between methodologies ranged from 0.76 to 0.90. These lower values are most likely due to differences in the treatment of time-dependent covariates.

The STM had greater predictive ability of daughter fertility at first insemination than the other methodologies. However, the CTM predicted daughter fertility more accurately in subsequent inseminations. The DPH and STM had a similar predictive ability of daughter fertility in second and subsequent inseminations.
(**Key words:** fertility, ordinal-censored threshold model, sequential threshold model, survival analysis)

Abbreviation key: CTM = ordinal-censored threshold model, **DPH** = discrete proportional hazard model, **INS** = inseminations per conception, **STM** = sequential threshold model.

INTRODUCTION

Fertility traits have been incorporated into national genetic evaluation systems of many leading dairy countries recently. Interval traits, such as days to first insemination, calving interval, or days open are most commonly used. However, the variation of these traits is highly dependent on management practices, such as estrus synchronization and differences in the voluntary waiting period (Wall et al., 2003). The number of inseminations to conception (**INS**) probably reflects actual variation of female fertility more closely, and it is one of the most important fertility traits from an economic point of view (González-Recio et al., 2004). The costs of semen, hormonal treatments, labor, and delayed subsequent calving increase rapidly as more inseminations are required for a cow to become pregnant. In addition, INS can reflect variation in both male and female fertility. An additional concern in the analysis of fertility is proper handling of cows that never become pregnant (i.e., censoring mechanisms). Furthermore, certain farmers may allocate cows to a natural service bull after several failed artificial inseminations and this leads to errors in services to conception data if such practice is not recorded. High quality reproductive data are needed to study this trait, but there are herds with missing or incomplete records in most dairy populations. Increasing the reliability of pregnancy check records could improve the quality of date from reproductive schemes. Therefore, it is necessary to develop suitable methods for analyzing INS data, to obtain accurate estimated breeding values and parameter estimates when using INS in a national genetic improvement program.

Three methods were developed and applied to field data on INS in Holstein cattle. First, an ordinal threshold model (Gianola, 1982; Gianola and Foulley, 1983) that accommodates censored records was implemented (**CTM**). Second, a sequential threshold model (**STM**), as described by Albert and Chib (2001), which can analyze categorical traits that occur in a sequential order, was applied. Third, a grouped survival model for discrete proportional hazard analysis (**DPH**), as developed by Prentice and Gloeckler (1978), was fitted. The STM

Received May 16, 2005.
Accepted June 23, 2005.
Corresponding author: Oscar González-Recio; e-mail: ogrecio@pan.etsia.upm.es.

and DPH models allow for time-dependent covariates, whereas CTM does not.

The objective of this research was to infer parameters of INS data with the aforementioned CTM, STM, and DPH models, and to assess their relative predictive abilities.

MATERIALS AND METHODS

Data

Data were provided by the regional Holstein Associations from the Basque and Navarra Autonomous Regions of Spain. Milk yield and reproductive data from 1994 through 2004 were used in the analysis. Records from embryo transfers were omitted, and cows needed to have a minimum of 100 DIM before culling and at least 1000 kg of total lactation milk yield to be included in the analysis. In addition, calving interval had to range between 300 and 600 d, and records were omitted if days to first service were unknown, less than 25 d, or greater than 160 d. Cows with a first calving before 18 mo or after 40 mo of age were excluded. At least 5 uncensored records were required per herd and per service sire, and herds with an average INS less than 1.5 were removed. A record was considered censored at a particular service if no subsequent calving was recorded, if the next breeding event was a natural service mating, or if no pregnancy was achieved after the fourth insemination. Four values (1, 2, 3, or 4) were possible for INS, and an indicator variable tagged each cow as being either pregnant or censored. Hence, cows with a censored record after 4 inseminations were included into a fifth category that represented more than 4 inseminations. The edited data set contained 80,071 lactation records and 143,927 insemination events from 47,509 cows. A total of 3267 bulls were present in the pedigree file.

Ordinal-Censored Threshold Model

The ordinal threshold model (Gianola, 1982; Gianola and Foulley, 1983) was extended to accommodate censored records. The ordinal-censored threshold model (CTM) postulates an underlying latent liability (λ) for number of inseminations to conception.

The statistical model for liability was:

$$\lambda_{jklmno} = \mathbf{x}'\beta_{jklmno} + \mathbf{h}_m + \mathbf{ss}_n + \mathbf{u}_o + \mathbf{e}_{jklmno}$$

The systematic effects ($\mathbf{x}'\beta$) in the model were as follows: days to first service treated as a covariate; effect of number of lactation (j = 1 to 4 levels); effect of calendar month of calving (k = 1 to 12 levels), and effect of year-season of calving (l = 1 to 30 levels). The random effects were: $\mathbf{h}_m$ = herd (m = 1 to 767 levels) distributed independently as $N(0,\mathbf{I}\sigma_h^2)$, where σ_h^2 is the variance among herds; $\mathbf{ss}_n$ = service sire for first insemination (n = 1 to 577 levels) distributed as $N(0,\mathbf{I}\sigma_{ss}^2)$, where σ_{ss}^2 is the variances among service sires; $\mathbf{u}_o$ = additive genetic effect of sire of cow (o = 1 to 3267 levels) distributed as $N(0,\mathbf{A}\sigma_u^2)$ where $\mathbf{A}$ is the additive relationship matrix between sires and σ_u^2 is the variances among sires of cows, and $\mathbf{e}_{jklmn}$ = random residual assumed independently distributed as $N(0,\mathbf{I}\sigma_e^2)$, where σ_e^2 is the residual variance, which was set equal to one. Service sires and sires of cows were assumed to be independently distributed.

When an observed response y_i falls into one of the possible known categories of INS, for example j, the liability of that observation is sampled from a truncated normal distribution between 2 given thresholds (T_{j-1} and T_j). Then, the conditional probability of the event can be written as:

$$\mathrm{Prob}(y_i = j, \text{ noncensored } |\beta, \mathbf{h}, \mathbf{ss}, \mathbf{u}, \mathbf{T}] =$$
$$\Phi[T_j - (\mathbf{x}_i'\beta + \mathbf{z}_{h,i}'\mathbf{h} + \mathbf{z}_{ss,j}'\mathbf{ss} + \mathbf{z}_{u,i}'\mathbf{u})]$$
$$- \Phi[T_{j-1} - (\mathbf{x}_i'\beta + \mathbf{z}_{h,i}'\mathbf{h} + \mathbf{z}_{ss,i}'\mathbf{ss} + \mathbf{z}_{u,i}'\mathbf{u})]$$

where j = 1, 2, 3, 4 indexes the category to which the uncensored observation y_i belongs; $\Phi(\cdot)$ is the standard normal distribution function; $\mathbf{x}_i$, $\mathbf{z}_{h,i}$, $\mathbf{z}_{ss,i}$, $\mathbf{z}_{u,i}$ are the respective incidence vectors of systematic (β), herd ($\mathbf{h}$), first service sire ($\mathbf{ss}$), and sire of cow ($\mathbf{u}$) effects, and $\mathbf{T} = [T_1, T_2, T_3, T_4]'$ is the vector of unknown threshold parameters. The thresholds must satisfy the restrictions $T_1 \leq T_2 \leq T_3 \leq T_4$, such that the cumulative distribution function is strictly nondecreasing. Further, the first threshold T_1 is set to zero, because this parameter cannot be identified in a probit analysis; hence, only T_2, T_3, T_4 are unknown.

If an observation is censored at the jth category, (e.g., a cow was not inseminated again beyond 2 services), then the liability is sampled from a left-truncated distribution. The truncation point is the threshold T_j corresponding to the last known insemination, and the probability can be written as:

$$\mathrm{Prob}(y_i = j, \text{ censored } |\beta, \mathbf{h}, \mathbf{ss}, \mathbf{u}, \mathbf{T}] =$$
$$1 - \Phi[T_j - (\mathbf{x}_i'\beta + \mathbf{z}_{h,i}'\mathbf{h} + \mathbf{z}_{ss,i}'\mathbf{ss} + \mathbf{z}_{u,i}'\mathbf{u})].$$

Then, assuming conditional independence, the joint distribution of the noncensored and censored observations can be written as