J. Dairy Sci. 88:2253–2259

Survival Analysis Applied to Genetic Evaluation for Female Fertility in Dairy Cattle

M. del P. Schneider,[1] E. Strandberg,[1] V. Ducrocq,[2] and A. Roth[3]
[1]Centre for Reproductive Biology in Uppsala and Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, PO Box 7023, SE-75007 Uppsala, Sweden
[2]Station de Génétique Quantitative et Appliquée, Institut National de la Recherche Agronomique, 78352 Jouy-en-Josas, France
[3]Swedish Dairy Association, Box 1146, SE-63180 Eskilstuna, Sweden

ABSTRACT

The objective of this research was to study whether survival analysis results in a more accurate genetic evaluation for female fertility traits compared with the usual methodology based on linear models. The fertility trait studied was interval between calving and last insemination. A stochastic simulation describing the reproductive cycle of first-parity cows was done, in which true breeding values for conception rate were created. A model containing effects of sire and herd was used both with survival analysis and with mixed linear model analysis to predict sire breeding values. Correlations between true breeding values for conception rate and breeding values for calving to last insemination predicted by the best survival analysis model or the best linear model were 0.77 and 0.68, respectively. The results showed that when pregnancy status is known, survival analysis is a better method than linear models for genetic evaluation of conception rate when using observations on the interval between calving and last insemination.
(**Key words:** female fertility, genetic evaluation, survival analysis)

Abbreviation key: CLI = interval between calving and last insemination, **CR** = conception rate, **PBV** = predicted breeding value, **TBV$_{\mathbf{CR}}$** = true breeding value for conception rate, **VWP** = voluntary waiting period.

INTRODUCTION

Poor reproductive performance is one of the most common reasons for culling in dairy herds (Dürr, 1997; Pryce et al., 1997; Swedish Dairy Association, 2002). The main costs associated with low fertility are higher insemination costs, lower production per day and, especially, higher replacement costs due to increased culling. Good female fertility is characterized by cows that return to cyclicity soon after calving, show strong signs of estrus, have a high probability of becoming pregnant when inseminated at the correct time, and have the ability to carry the resulting fetus to term. Among the potential measures that can be used to describe this complex trait, this study emphasizes the interval between calving and last insemination (**CLI**, also called days open). The trait CLI is a measure that is a combination of return to cyclicity, expression of estrus, and ability to conceive (conception rate). If insemination dates are available, CLI can be used in breeding programs, which is the case in some countries (Mark et al., 2001).

With field data, the pregnancy status of cows is not always available and thus one cannot be sure if cows have conceived (Weller and Ron, 1992; Roxström, 2001). However, even if pregnancy information is available, linear model methodology, the method most frequently used for the genetic evaluation of fertility, has the disadvantage that it cannot properly distinguish between pregnant and nonpregnant cows. Hence, records of pregnant and nonpregnant cows have to be treated alike (as is commonly done for interval from calving to last insemination), or the records of nonpregnant cows have to be excluded (as is commonly done for calving interval) or extended by projection. Culling for reproduction creates another problem. The worse a bull's daughter fertility is, the larger the proportion of daughters culled for reproductive failure. Thus, sires are evaluated without correct information on their daughters with poor fertility (these daughters either have missing information or observed intervals that are shorter than true intervals). Therefore, such bulls appear to be better than they really are and this is expected to lead to less efficient selection.

Survival analysis is an alternative method for analyzing reproductive traits recorded as time intervals (Lee et al., 1989; Eicker et al., 1996; Harman et al., 1996; Allore et al., 2001). Survival analysis is a statistical

---

Received June 8, 2004.
Accepted March 20, 2005.
Corresponding author: Erling Strandberg; e-mail: Erling.Strandberg@hgen.slu.se.

method for studying the occurrence and timing of events, where the outcome variable corresponds to a measure of time elapsed from a starting point until the occurrence of a certain event (Lee, 1992). The length of this interval is not always known, because competing events may occur before the occurrence of the event under study. For example, in our case, cows may have been culled, sold, or the study may have stopped before the cows conceived. One of the main advantages of survival analysis is that it can retain the information from cows that are culled before conception or not pregnant by the time the data recording was completed. Thus, records from pregnant (uncensored) and nonpregnant (censored) cows can be treated jointly and included in the analysis, making proper use of all the available information. Within the field of fertility in dairy cattle, survival analysis has been applied to study: 1) the effects of diseases on days to conception (Lee et al., 1989; Harman et al., 1996b), 2) the relationship between BCS and postpartum reproductive efficiency (Suriyasathaporn et al., 1998), and 3) the effect of early lactation milk yield on days open (Harman et al., 1996a). So far, no research using genetic models with survival analysis has been published for fertility traits.

The objective of this study was to investigate by simulation whether the analysis of CLI using survival analysis results in a more accurate genetic evaluation for conception rate than do the commonly used approaches based on linear models.

MATERIALS AND METHODS

A simulation was done to create phenotypic observations for CLI, some of which were censored observations, that is, cows that did not get pregnant and were culled. To avoid the possibility that the simulation itself would favor any of the ensuing statistical analysis methods, we did not simulate breeding values directly for CLI. Rather, we created 3 underlying traits: milk production, interval between calving and first ovulation, and conception rate. Then we added the effect of decision-making of farmers, such as number of inseminations allowed and voluntary waiting period. We then simulated the reproductive performance of each individual and ended up with the trait that we were interested in studying: CLI.

Simulated Data

Each replicate of the simulated data consisted of 60,000 first-parity cows, daughters of 400 unrelated sires distributed over 1200 herds. The herd size was fixed to 50 cows. The average number of daughters per sire was 150 (SD = 12.3), ranging from 104 to 201 daughters. Fifty replicates were done. Three traits were simulated: 305-d milk production (kg), interval between calving and first ovulation (d), and conception rate (**CR**, %). The mean phenotypic values were 8000 kg (SD 1000) and 28 d (SD 15) for milk production and interval between calving and first ovulation, respectively.

Table 1. Heritability and genetic correlations assumed for the simulated traits (diagonal = heritability; below diagonal = genetic correlation; above diagonal = environmental correlation).

	305-d milk production	CFO	Conception rate
305-d milk production	**0.30**	0.00	0.00
CFO	0.10	**0.20**	0.00
Conception rate	–0.10	0.00	**0.05**

CFO = Interval between calving and first ovulation.

Conception rate was simulated as a binary trait with an underlying normally distributed liability for conception with mean zero and standard deviation of unity [~N(0,1)]. Zero was chosen as the threshold; hence, all phenotypic values above 0 corresponded to pregnant cows (50% CR). Heritabilities and genetic and environmental correlations among the traits are shown in Table 1. The heritability of the interval between calving and first ovulation was chosen according to the estimates for the interval from calving to commencement of luteal activity reported by Darwash et al. (1997), Veerkamp et al. (1997), and Royal et al. (2002). For CR, we assumed a heritability value somewhat higher than the values found in the literature, which were estimated with linear model methods, because we simulated CR on the underlying scale. Herd variances as proportion of the phenotypic variance were 9% for the 3 traits. The phenotypic value for each trait was created as: phenotypic value = mean + herd effect + breeding value (½ sire breeding value + ½ dam breeding value + Mendelian sampling) + environmental value.

Simulation Process

The simulation of the reproductive cycle is summarized in Figure 1. For all herds, the voluntary waiting period (**VWP**; the first part of the lactation during which no insemination occurs) was set to 8 wk and the heat detection rate to 60%. Heat detection was drawn from a uniform distribution between 0 and 1; if the value was below 0.6, the heat was assumed to be detected. For each herd a maximum number of inseminations (MAXHERD) was set as the integer part of a random draw from the distribution ~N(5,1) Within herd, the maximum number of inseminations (MAXINS) for each cow was calculated according to its milk yield deviation from herdmates as the integer part of (MAXHERD) +