Fertility, Mastitis and Longevity in Dairy Cattle Analyzed Using Survival Models

Maria del Pilar Schneider
Faculty of Veterinary Medicine and Animal Science
Department of Animal Breeding and Genetics
Uppsala

Doctoral thesis
Swedish University of Agricultural Sciences
Uppsala 2006

Acta Universitatis Agriculturae Sueciae
2006: 45

ISSN 1652-6880
ISBN 91-576-7094-3

Tryck: SLU Service/Repro, Uppsala 2006

Abstract

Schneider, M. del P. 2006. Fertility, Mastitis and Longevity in Dairy Cattle Analyzed Using Survival Models. Doctoral dissertation.
ISSN 1652-6880, ISBN 91-576-7094-3

The aim of the thesis was to investigate whether survival analysis (SA) results in a better genetic evaluation of female fertility and mastitis traits and to study the effects of mastitis and pregnancy status as risk factors for culling. Sire breeding values for interval between calving and last insemination (CLI) and interval between first and last insemination (FLI) were predicted using SA and mixed linear models (LM). Correlations between simulated true breeding values for conception rate and breeding values for CLI and FLI predicted by SA were higher than corresponding correlations with LM. When pregnancy status was known, SA was better than LM for genetic evaluation of conception rate when using observations on CLI and FLI. If selection were carried out on these predicted breeding values, this would translate into 8 to 12% higher genetic progress for FLI and CLI, respectively.

Clinical mastitis was analyzed with LM (binary trait), and time to first mastitis with SA. The higher accuracies for SA could be translated into a higher genetic response, approximately, 3% for first parity and 25 % for later parities.

The effect of pregnancy status and mastitis on culling in Swedish dairy cattle was analyzed with SA. Mastitis affected culling decisions throughout the lactation, but its effect depended on pregnancy status and the stage during which the cow was treated. The risk of culling was low for pregnant cows, whether or not they had been treated for mastitis. Different patterns were observed for the risk of culling between open and pregnant cows. For both groups, the risk of being culled was higher for cows treated for mastitis than for untreated cows. For open cows, the risk was higher for cows treated in earlier stages, whereas for pregnant cows, the risk was similar across different stages of lactation.

This thesis showed the potential of using SA for genetic evaluation of fertility traits and clinical mastitis in dairy cattle. SA was also found to be a useful method to analyze the effect of pregnancy status and clinical mastitis as risk factors for culling by treating them as time-dependent covariates.

Keywords: female fertility traits, mastitis, pregnancy status, longevity, survival analysis, mixed linear models

Author's address: Maria del Pilar Schneider, Department of Animal Breeding and Genetics, P.O. Box 7023, Swedish University of Agricultural Sciences, S-750 07, Uppsala, Sweden.
Email: Pilar.Schneider@hgen.slu.se

Contents

Appendix

Papers I - IV

The present thesis is based on the following papers, which will be referred by their Roman numerals:

I. Schneider, M. del P., Strandberg, E., Ducrocq, V. and Roth, A. 2005. Survival analysis applied to genetic evaluation for female fertility in dairy cattle. *Journal of Dairy Science* 88, 2253-2259.

II. Schneider, M. del P., Strandberg, E., Ducrocq, V. and Roth, A. 2006. Short Communication: Genetic evaluation of interval from first to last insemination with survival analysis and linear models. (Submitted).

III. Carlén, E., Schneider, M. del P. and Strandberg, E. 2004. Comparison of between linear models and survival analysis for genetic evaluation of clinical mastitis in dairy cattle. *Journal of Dairy Science* 88, 797-803.

IV. Schneider, M. del P., Strandberg, E., Emanuelson, U., Grandinson, K., Roth A. 2006. The effect of mastitis and pregnancy status on culling in Swedish dairy cows. (Submitted).

Abbreviations

CLI: interval from calving to last insemination
FLI: interval from first to last insemination
LM: mixed linear models
MAST: mastitis (healthy/sick)
PBV: sire predicted breeding value
SA: survival analysis
SCC: somatic cell count
SLB: Swedish Friesian
SRB: Swedish Red and White
TBV_{CR}: sire true breeding value
TFM: time to first mastitis
TH: threshold models

Introduction

Fertility, mastitis resistance and longevity are functional traits which share some common features. They can be measured as failure time traits and they are subject to censoring. When fertility and mastitis are studied as risk factors for culling, they can be treated as time-dependent covariates. Failure time traits can be expressed as a measure of time elapsed from a starting point (origin) to an end point when a certain event might occur (sometimes called failure). For fertility, the time interval can be measured from calving until conception; for mastitis, from calving to the outbreak of the disease; and for longevity from first calving to culling.

The length of these time intervals is not always known: for some individuals the event of interest is not observed (cows still not pregnant, still healthy or still alive). Moreover, competing events may occur before the occurrence of the event under study. For example, when studying fertility, some cows are culled before they have the opportunity to conceive. Both situations lead to censored observations.

Culling, conception and outbreak of mastitis can depend on several covariates. Some covariates, like age at first calving and breed, remain constant over time. Others, like year and season, change their value during the period under study. These covariates are called time-dependent. If mastitis and pregnancy status are to be studied as risk factors for culling, they can be defined as time-dependent covariates. Mastitis can occur at any time during the lactation, and this can be accounted for in the analysis of longevity by considering different stages of disease occurrence (Rajala-Schultz & Gröhn, 1999a). The same reasoning is valid when studying the effect of pregnancy status; cows can get pregnant at different stages during the lactation. The timing of these effects may influence culling decisions.

Some problems and drawbacks of the current approach to analyze fertility and mastitis

Different fertility traits, ranging from binary (discrete) responses to continuous (or interval) variables, can be used for genetic evaluation (Jorjani, 2005). Two such potential continuous traits are the interval between calving and last insemination (CLI), also called days open, and the interval between first and last insemination (FLI). CLI is a measure of both the return to cyclicity after calving and the ability to conceive (conception rate), whereas FLI is mainly a measure of conception rate.

Analysis of these traits has a problem: how to handle records of non-pregnant cows (censoring). In fertility studies there are cows which are not pregnant at the end of the period analyzed and cows that are culled for reproductive problems. Mixed linear models (LM) are currently used to predict breeding values for fertility traits (Interbull, 2006). However, LM is not well suited for handling censored observations. The practical solutions include: 1) to handle records from pregnant and non-pregnant cows in the same way; as is commonly done for CLI,

2) to exclude records of non-pregnant cows; as is usually done for calving interval, and 3) to extend records by projection.

A large part of the non-pregnant cows are actually culled for reproductive reasons, *i.e.* they did not get pregnant within the period allowed. Culling for reproduction creates another problem, *i.e.*, sires with low genetic merit for daughter fertility have a larger proportion of daughters culled. If culling is not considered properly, sires are evaluated with the wrong information on their daughters with poor fertility, *i.e.* daughters that either have missing information or observed intervals are shorter than the true ones. Therefore, the hypothesis is that bulls would appear to be better than they are, which in turn probably leads to less efficient selection.

In the genetic evaluation of clinical mastitis, cross-sectional LM are applied to predict breeding values. Mastitis is commonly defined as an all-or-none trait, healthy or sick, within a defined period of the lactation (Interbull, 2006). However, this definition ignores the timing of mastitis and it is thus not possible to differentiate between cows getting mastitis in early or late lactation. Furthermore, cows culled before they express mastitis have to be treated as either missing or healthy, which may introduce some bias when culling is related to mastitis. Consequently information is lost when mastitis is defined as an all-or-none trait.

The advantage of survival analysis to analyze fertility, mastitis, and risk factors for culling

Survival analysis (SA) is a statistical method for studying the occurrence and timing of events (Lee, 1992). It can handle both censoring and time-dependent covariates. In dairy cattle breeding SA is applied for routine genetic evaluation for longevity in many countries (Interbull, 2006).

The advantage of applying SA to study fertility traits is that information from cows that are not pregnant or culled before conception is retained. Thus, records from pregnant (uncensored) and non-pregnant cows (censored) can be treated jointly and included in the analysis making proper use of all the available information. SA has been used to study 1) the effects of diseases on days to conception (Lee *et al.*, 1989; Andersson *et al.*, 1991; Harman *et al.*, 1996b), 2) the relationship between body condition score and postpartum reproductive efficiency (Suriyasathaporn *et al.*, 1998), and 3) the effect of early lactation milk yield on days open (Harman *et al.*, 1996a). Limited research using genetic SA models has been done for fertility traits. Weigel (2004) analyzed two fertility traits in North American herds using survival models: days from calving until first positive pregnancy examination, with a Weibull model, and services to conception, with a discrete time model. Recently, González-Recio *et al.* (2005) studied the number of inseminations to conception in Holstein cows using a grouped survival model. Alternative approaches to SA also have been proposed: a longitudinal Bayesian threshold analysis (Averill *et al.*, 2004) and a bivariate censored threshold-linear model (Chang *et al.*, 2006).

A trait such as time to first mastitis can be used to analyze clinical mastitis with SA. This trait definition has three advantages. First, the timing of mastitis is considered (the observation becomes uncensored when the cow is treated). Second, observations for healthy cows are treated as censored. Third, records of cows culled for other reasons (possibly correlated to mastitis) before they have had the full opportunity to express mastitis also become censored. Culled cows are considered as healthy but only for the period they remained in the herd, whereas with the traditional approach these cows are treated as completely healthy, which might introduce some bias.

Genetic studies of mastitis using SA, or similar methodologies, have so far been limited. Time to first veterinary treatment of clinical mastitis had been analyzed with: 1) a stochastic process model and a semi-parametric proportional hazards model (Saebø *et al*., 2002), 2) a Bayesian proportional hazards model, including both genetic and environmental covariates (Saebø & Frigessi, 2004), and 3) a competing process model (Saebø *et al*., 2005). Although survival models have been developed specifically to handle failure time traits, other approaches have been applied to analyze mastitis: cross-sectional threshold models (Heringstad *et al*., 2001), longitudinal threshold models (Heringstad *et al*., 2003; Chang *et al*., 2004), and multivariate threshold models (Heringstad *et al*., 2004).

When studying the effect of mastitis as a risk factor for culling, the timing of the disease, *i.e.* when mastitis occurs during lactation, and when culling occurs should be considered (Gröhn *et al*., 1997). The same situation applies when studying the effect of pregnancy status on culling. When the effect of a disease is modelled as a time-independent covariate, its effect on the outcome is assumed to be the same both before and after the occurrence of the event, which only makes sense when the disease occurs early in lactation. In contrast, when the disease is considered as a time-dependent covariate, its effect differs before and after the occurrence (Beaudeau *et al*., 2000). SA can handle time-dependent covariates, and thus the timing of diseases can be considered; whereas this is not possible to address when standard regression techniques (*e.g.,* logistic regression) are used.

Research has shown that both mastitis and poor fertility substantially affect culling rates. Mastitic cows have a higher risk of being culled than healthy cows (Beaudeau *et al.*, 1995; Gröhn *et al.*, 1998; Rajala-Schultz & Gröhn, 1999b; Neerhof *et al.*, 2000). A relationship between the time mastitis occurs and time of culling has been also reported (Gröhn *et al.*, 1997, 1998; Rajala-Schultz & Gröhn, 1999a). Beaudeau *et al.* (1995) reported that cows with poor reproductive performance have an increased risk of being culled. Gröhn *et al.* (1998) found that once a cow had conceived, her risk of being culled dropped sharply. Similar results about the effect of mastitis and pregnancy status on culling were obtained in Swedish dairy cattle (Schneider *et al.*, 2005). However, in these studies an interaction between reproductive performance and incidence of mastitis was not considered. Therefore it was not possible, for example, to differentiate between the risk of being culled of a pregnant and open (non-pregnant) cow that had been treated for clinical mastitis.

Importance of fertility, mastitis, and risk factors for culling

Poor reproductive performance and mastitis, including high somatic cell count (SCC), are the main reasons for culling in Swedish dairy cattle (Swedish Dairy Association, 2000). Each of them account for almost 1/4 of the total culling. The main costs associated with low fertility are higher insemination costs, lower production per day and, especially, higher replacement costs owing to increased culling. Mastitis negatively affects dairy farm profitability through an increase of veterinary and treatment costs, reduced milk yield, discarded milk, reduction in milk price, and increased culling. Furthermore, aspects regarding animal welfare and food security have to be considered. The incidence of impaired fertility and mastitis can be reduced, both by means of genetic improvement, and by controlling the environment through better management practices. Selection measures give slower effects, but they accumulate over generations, and breeding can thus be considered a permanent and cost-efficient method (Strandberg & Shook, 1989).

In dairy cattle selection, the main emphasis has been placed on production traits. Except for the Nordic countries, fertility and health were seldom considered, and due to their antagonistic genetic relationship with production traits (Roxström, 2001; Carlén, 2003; VanRaden *et al.*, 2004), breeding has lead to a deterioration in functional traits. Therefore, fertility traits and mastitis resistance are increasingly being incorporated in breeding programs in countries outside Scandinavia. Besides the mentioned problems regarding censoring, there are additional issues related to these traits: the heritability is low, ranging from 1 to 10% (Mark *et al.,* 2001), they are difficult and expensive to record, and some traits may be strongly affected by management decisions. Accurate genetic evaluations are needed to identify the best animals in the population. When better methods are applied to predict breeding values, higher genetic response can be expected.

To improve the environment by applying better management practices, it is important to better understand how fertility and mastitis affect the culling of cows. If mastitis and failure to conceive considerably affect the risk of culling at different stages of the lactation, environmental and management measures should be focused on these periods.

Aims of the thesis

- To study whether the analysis of fertility traits (interval from calving to last insemination and interval from first to last insemination) and time to first mastitis using survival analysis results in a better genetic evaluation than the currently widely used methods.

- To estimate the effects of environmental factors, mastitis and pregnancy status on longevity.

Overview of the investigations

Materials

In Paper I and II a stochastic simulation was used to create phenotypic observations for CLI and FLI. Three underlying traits were simulated: 305-d milk production, interval from calving to first ovulation and conception rate. The effect of decision making of farmers, such as number of inseminations allowed and voluntary waiting period, was also considered. Thus, the reproductive performance of each individual was simulated and phenotypic observations for CLI and FLI were obtained. With this approach, true breeding values for conception rate (TBV_{CR}) were known. Each replicate consisted of 60 000 first parity cows, daughters of 400 unrelated sires distributed over 1200 herds. The average number of daughters per sire was 150 (SD 12.3), ranging from 104 to 201 daughters. For FLI, a smaller progeny group was also studied. Herd size was fixed at 20 resulting in an average daughter group size of 60, varying between 38 and 84 (SD 7.8), and a total number of 24 000 cows.

CLI was defined as the interval from calving to conception (at last known insemination) or censoring (at last known insemination). For a cow that was never detected in heat, and thus never inseminated (approximately 8% of the records), the record was censored at the maximum waiting period assigned for that cow. FLI was defined as the interval from first insemination to conception (at last known insemination) or censoring (at last known insemination). For a cow that was never detected in heat, and thus never inseminated (approximately 8% of the records), FLI was equal to the maximum allowed insemination period assigned for that cow.

The same trait definitions were used for LM and SA. In SA, records of pregnant cows were considered as uncensored, and non-pregnant cows as censored. In the LM analysis, records of pregnant cows were not distinguished from non-pregnant cows. In Figure 1 the trait definition and censoring are exemplified for CLI. Cow 1 conceived on day 119 after calving (at the 3rd insemination): her observation was treated as uncensored. Cow 2 did not conceive at the last insemination, which occurred on day 161 and she was censored on that day.

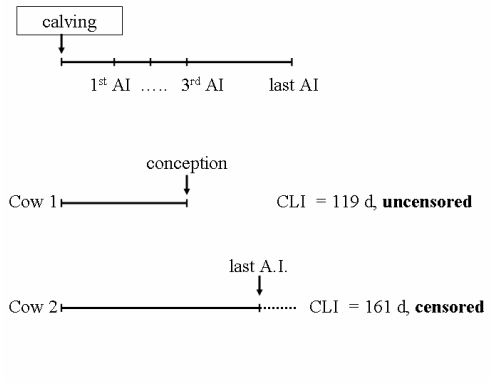

Fig. 1. Example of trait definition and censoring for the interval between calving and last insemination (CLI). AI: artificial insemination.

Field data for Paper III and IV was provided by the Swedish Dairy Association. Information on individual cows about production, reproduction, veterinary treatments and reasons for culling was available. Under Swedish regulations, treatment for clinical mastitis and other diseases must be administered by veterinarians and recorded in the official health-recording system.

Paper III included data from the first three lactations of Swedish Holstein cows having their first calving between 1995 and 2000. The number of records analyzed was 221 104, 122 280 and 59 233 for lactation 1, 2 and 3, respectively. The pedigree file had 1139 bulls including 838, 784 and 673 sires with daughter records for lactation 1, 2 and 3, respectively. Sires with less than 50 daughters were excluded. For the LM analysis, clinical mastitis was defined as a binary trait: MAST=1 when a cow had a veterinary-treated clinical mastitis (with or without teat injury) or was culled due to mastitis from day 10 before calving to day 150 after calving, and MAST=0 otherwise. Figure 2 illustrates MAST. Cow 1 had a mastitis treatment during the period analyzed: she was considered as sick (MAST=1). Cow 2, which was culled 90 days after calving, and cow 3, which did not have mastitis until day 150, were both treated as healthy (MAST=0).

In SA, for a sick cow time to first mastitis (TFM) was defined as the number of days from day 10 before calving to the day of the first treatment of mastitis or culling due to mastitis (uncensored observation). For a healthy cow, TFM was defined as the number of days between day 10 before calving until: 1) the day of next calving, 2) the day of culling for another reason than mastitis, 3) the day of movement to another herd, or 4) lactation day 240 (censored observation).

Figure 2 shows an example of TFM: cow 1 was treated for mastitis on day 10 after calving: her record was uncensored on day 20. Cow 2 was culled on day 80 after calving and she never got sick during that period: she was censored on day 90. Cow 3 never had mastitis during the lactation: she was censored on day 370 when she had a new calving.

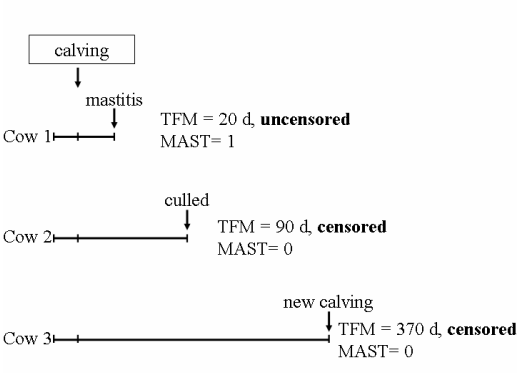

Fig. 2. Example of trait definition for mastitis (MAST) and time to first mastitis (TFM).

In Paper IV, data on 980 705 cows calving for the first time between January 1 1988 and December 31 1996 and distributed across 15 408 herds were analyzed. Four breeds were included: Swedish Red and White (SRB), Swedish Friesian (SLB), Swedish Polled Breed and Jersey, as well as crossbred cows (SRB x SLB). Productive life was defined as the number of days between first calving and culling or censoring. Records were defined as censored when cows were still alive at the end of the study period (*i.e.* on 31 December 1996) and uncensored otherwise.

Methods

SA (Paper I to IV) and LM (Paper I to III) were applied for data analysis. The software packages used were Survival Kit V3.12 (Ducrocq & Sölkner, 1998) and DMU (Jensen & Madsen, 1994), respectively. In DMU, variance components were estimated using the average information residual (restricted) maximum likelihood (REML) algorithm based on mixed linear models (Jensen *et al.*, 1997). The models supported by the Survival Kit belong to the proportional hazards models with a single continuous or discrete response time:

$$\lambda(t, \mathbf{x}(t), \mathbf{z}(t)) = \lambda_0(t) \exp\left\{\mathbf{x}(t)'\mathbf{b} + \mathbf{z}(\mathrm{t})'\mathbf{u}\right\}$$

where, $\lambda(t, \mathbf{x}(t), \mathbf{z}(t))$ is the hazard function of an individual depending on time *t*; $\lambda_0(t)$ is the baseline hazard function where t is continuous, $\mathbf{x}(t)$ is a vector of (possibly) time-dependent fixed covariates with corresponding parameter vector **b,** and $\mathbf{z}(t)$ is a vector of (possibly) time-dependent random covariates with corresponding parameter vector **u**.

The baseline hazard function can be either unspecified (semi parametric) or can have a parametric form (*e.g.*, Weibull). In the first case the baseline hazard function is left completely arbitrary. This results in a semi parametric regression model known as the Cox model (Cox, 1972). This model permits the estimation of parameters without making any assumption about the distributional form of $\lambda_0(t)$. In the second case, when a Weibull hazard distribution is assumed for the baseline hazard function, it defines a Weibull model, where $\lambda_0(t) = \lambda\rho(\lambda t)^{\rho-1}$ with positive scale parameter λ and positive shape parameter ρ. When observations are expressed on a discrete scale, $\lambda_0(t)$ is a piecewise constant function which describes the discrete hazard for each particular interval and this results in the grouped data model (Prentice & Gloeckler, 1978).

The linear and survival models in Paper I (CLI) and II (FLI) included the random effects of sire and herd. For SA, for both traits, the following models were used: Cox (S1), Weibull (S2), and a Weibull model (S4) in which all the records were considered as uncensored to evaluate the effect of proportional hazards models to handle censoring. Two additional models were also studied: for CLI, a Weibull model (S3_I) with the origin shifted to avoid a long early period without events; for FLI, a grouped data model (S3_II) in which FLI was divided into 9 periods (discrete scale): 1 day; 2-31 days; 32-52 days; 53-73 days; 74-94 days; 95-115 days; 116-136 days; 137-157 days; and ≥ 158 days after first

insemination, respectively. Each period approximately represents an insemination number.

For the LM analysis two models were applied: L1, to analyze CLI or FLI, and L2, to analyze the log transformation of the traits instead. For CLI, an extra model from which non-pregnant cows were excluded was also considered (L3).

Variance components and sire predicted breeding values (PBV) were estimated. The model comparison criterion was Pearson correlations (SAS Institute, 1999) between sire PBV and true breeding values for conception rate (TBV_{CR}).

The linear and survival (Weibull proportional hazards) models used in Paper III to analyze mastitis included the fixed time-independent effects of year by month of calving, age at calving, proportion of heterosis (regression), proportion of North American Holstein genes (regression) and the random effects of herd by year of calving and sire.

Variance components and sire predicted breeding values were estimated. Heritabilities and accuracies of selection (r_{TI}) were calculated for model comparison. The heritability for the LM was calculated as $h^2 = 4\sigma_s^2 \Big/ \left(\sigma_s^2 + \sigma_e^2\right)$, where σ_s^2 is the sire variance, and σ_e^2 is the residual variance. For SA, the heritability was defined as:

$$h_{equ}^2 = 4\sigma_s^2 /(\sigma_s^2 + 1/p)$$

where p is the proportion of uncensored records. This derivation for the heritability on the original scale, which is not dependent on the Weibull parameters, was suggested by Yazdi *et al.* (2002) as the equivalent heritability. The authors showed very good agreement between accuracy and selection response calculated using h_{equ}^2 and observed accuracies and selection responses calculated from simulation. The term equivalent refers to the fact that the predicted breeding value of a sire with n daughters would get the same reliability as if it were evaluated on a linear trait with this heritability. An increase in the proportion of uncensored records with time implies that the equivalent heritability increases with time until it reaches the theoretical heritability [$h^2 = 4\sigma_s^2 /(\sigma_s^2 + 1)$] that would be obtained in total absence of censoring.

The accuracy of selection (r_{TI}) was calculated as $r_{TI} = \sqrt{n/(n+k)}$, where n is the number of daughters and $k = (4 - h^2)/h^2$ or $k = (4 - h_{equ}^2)/h_{equ}^2$.

A Weibull proportional hazards model was used to analyze the effect of pregnancy status and mastitis on culling (Paper IV). The model included the fixed time-dependent effects of parity by pregnancy status by mastitis, year by season, and peak test day yield deviation. It also included the time-independent fixed effects of age at first calving, breed, region, herd production level and the random effect of herd.

The time-dependent effect of pregnancy status was defined as $i \cdot j$, where i = 1-5, indicating the lactation stage at risk of culling (0-60, 61-150, 151-240, 241-305, and > 305 days after calving), and j = 0-5, indicating lactation stage when the cow was assumed to have become pregnant (0 indicating a non-pregnant cow). As an example, a cow that did not get pregnant would have a sequence 1·0, 2·0, 3·0, 4·0, 5·0, whereas a cow that became pregnant in stage 3 would have a sequence 1·0, 2·0, 3·3, 4·3, 5·3.

The time-dependent effect of mastitis was defined similarly as $k \cdot l$, where k = 1-5, indicating the lactation stage at risk of culling (0-30, 31-60, 61-150, 151-240, and > 240 days after calving), and l = 0-5, indicating lactation stage when the cow was first treated for mastitis (0 indicating a non-treated cow). For example, a non-treated cow would have the sequence 1·0, 2·0, 3·0, 4·0, 5·0, whereas a cow treated for mastitis in stage 2 would have a sequence 1·0, 2·2, 3·2, 4·2, 5·2.

The classification of pregnancy status and mastitis gave a total of 82 combinations for each parity group (parity 1 and ≥ 2). In this way classes for open and pregnant cows, pregnant at different stages of lactation, were combined with classes of healthy and treated cows, treated at different stages of the lactation (for more details see Figure 3 in Paper IV).

Results

Fertility (Paper I, II)

Correlations between sire TBV_{CR} and PBVs for both traits, CLI (Paper I) and FLI (Paper II) are summarized in Table 1. For both traits, correlations between TBV_{CR} and sire breeding values predicted by SA models S1-S3 were higher than the corresponding correlations from LM. For CLI, the Cox model (S1) and the Weibull origin shifted model (S3_I) had the highest correlations with TBV_{CR}. For FLI, PBVs from the grouped data model (S3_II), in which FLI was expressed in nine periods, and the Weibull model (S2), had the highest correlations with TBV_{CR}.

Among the LM the log transformation of CLI or FLI (model L2) had the highest correlation between TBV_{CR} and PBV. In SA when all data were treated as uncensored (model S4), the correlation was similar to those from the LM.

Table 1. *Correlations between true breeding values for conception rate and predicted breeding values from survival analysis and linear models (r) and accuracies (r_{TI}) calculated from the estimated heritabilities and number of daughters for CLI and FLI (Mean and standard error based on 50 replicates) (from Paper I and II)*

	CLI		FLI	
	r	r_{TI}[1]	r	r_{TI}[1]
Survival analysis				
S1: Cox	$0.767_{.003}$	0.77	$0.791_{.003}$	0.67
S2: Weibull	$0.747_{.003}$	0.82	$0.799_{.003}$	0.81
S3_I: Weibull origin shifted	$0.769_{.003}$	0.76	-	-
S3_II: Grouped data	-	-	$0.803_{.003}$	0.80
S4: Weibull uncensored	$0.665_{.004}$	0.73	$0.737_{.003}$	0.77
Linear models				
L1: All records	$0.677_{.004}$	0.78	$0.716_{.004}$	0.80
L2: Log transformed	$0.685_{.004}$	0.75	$0.744_{.003}$	0.76
L3: Non-pregnant cows excluded	$0.516_{.006}$	0.69	-	-

[1] Daughter group size is assumed to be 150.

For FLI, all models were also analyzed using a smaller progeny group size of 60. Although the correlations between TBV_{CR} and PBVs were smaller compared with the correlations obtained with the larger progeny group (150 daughters per sire), the same pattern was observed; breeding values predicted by SA had higher correlations with TBV_{CR} than did PBVs from LM (SA models S1 to S3_II: 0.632, 0.644, 0.646 *vs.* LM models L1 and L2: 0.580 and 0.586, respectively).

Mastitis (Paper III)

Estimates of heritabilities and accuracies from LM (MAST) and SA (TFM) are presented in Table 2. The heritability estimates from SA were higher than the heritability estimates from LM in all parities. The corresponding accuracies in selection were also higher for SA, only slightly in first parity while considerably higher in second and third parity.

Table 2. *Heritabilities and accuracies (calculated from the estimated heritabilities and number of daughters) of clinical mastitis analyzed by linear models (LM) and survival analysis (SA) for the first three lactations of Swedish Holstein cows (from Paper III)*

	Heritability		Accuracy (r_{TI})[1]	
Parity	LM	SA	LM	SA
1	0.032	0.036	0.74	0.76
2	0.014	0.030	0.54	0.68
3	0.014	0.027	0.48	0.60

[1] Daughter group size is assumed to be 150 for first parity, with 75% survival to next lactation.

Effects of pregnancy status and mastitis on culling (Paper IV)

Pregnancy status and mastitis treatment affected culling decisions and an interaction between these two effects was found. The farmer's knowledge of pregnancy status dramatically affected culling: the risk of culling of pregnant cows was sharply reduced, irrespective of whether the cow was healthy or not. However, the later in the lactation a cow became pregnant, the greater the risk of being culled.

Cows treated for mastitis were at higher risk of being culled than healthy animals. The risk of being culled was two times higher for a cow treated in the first 30 days after calving than it was for a healthy animal. Mastitis affected culling decisions throughout lactation, but its effect depended on pregnancy status and on the stage at which the cow was treated. For open cows mastitis affected culling decisions more in early lactation. For pregnant cows, mastitis affected culling decisions across the different stages of lactation similarly, but it did not immediately affect the culling risk for pregnant cows in stage 61-150 days. The risk of being culled was markedly smaller for cows treated at late stage during the lactation (> 240 days) whether pregnant or not. The same trends were found for parity 1 and later parities (≥ 2).

General discussion

Fertility (Paper I, II)

SA gives better prediction of breeding values than LM

As shown in the results, SA is better than LM at predicting the genetic merit of bulls for conception rate when using observations on FLI or CLI. If selection is carried out on these PBVs, this would also translate into 8 to 12% higher genetic progress for FLI and CLI, respectively.

A basic Weibull model works unexpectedly well

For CLI, Weibull model S3_I was better than the basic Weibull model S2, mainly because it is difficult for S2 to handle the zero hazard before day 56. In the simulation, a voluntary waiting period of 56 days during which no conception occurred was assumed. Model S3_I with a suitable origin shift, excluding the time where no failures occur, performed as well as the more computationally demanding Cox model (S1).

For FLI, the grouped data model (S3_II) had a slightly better correlation than the basic Weibull model (S2). Model S2 was not expected to be the best model because the assumption of the Weibull distribution did not hold well at the beginning of the period analyzed (Figure 2 in Paper II). Nevertheless, S2 was quite robust to correctly predict breeding values given the distribution of FLI and the deviation from the Weibull assumption. The advantage of the grouped data model is that it requires no particular assumption about the shape of the baseline distribution and overcomes the problem of many ties (equal failure times). Thus, the grouped data model was expected to be the most suitable to estimate breeding values when analyzing FLI.

The Cox model (S1) had a good fit for CLI, but not for FLI. The advantage of this model is that no assumption is made about the form of the baseline hazard function. For CLI, S1 seems to fit the data better than the Weibull model S2, especially at the beginning of the period analyzed where no failures occur. For FLI a better fit was expected; however, in the presence of many ties this model has a poor fit. FLI had a special distribution (Figure 1 in Paper II) where 50% of the cows conceived at the first insemination and the other cows potentially conceived at intervals of 21 days on average, a situation which creates many ties.

Correct handling of censoring records is important

When all records were treated as uncensored in SA (S4), the correlations between TBV_{CR} and PBV were similar to the LM. These results showed how the proportional hazards models can handle censoring, which is one of the advantages of SA. It makes proper use of information that would be otherwise discarded (exclusion of non-pregnant cows) or treated as uncensored (no distinction between pregnant and open cows). Carriquiry *et al.* (1987) concluded that censoring

produces a general loss of information; when this is ignored in the statistical analysis, inferences may be distorted. The severity of the problem depends on the underlying distribution and on the extent of censoring.

Log transformation gives a slight advantage

In the LM analysis, the log transformation of the trait (L2) gave a small improvement; correlations between TBV_{CR} and PBV were higher than corresponding correlations using the original scale (L1). The distribution of CLI (Figure 2 in Paper I) is skewed and not normal. The log transformation of the data cannot solve this problem. VanRaden (2003) reported that the log transformation of number of lactations, used as dependent variable in an analysis of longevity, resulted in lower heritability and accuracy as compared with the original scale. The author stated that log transformation can be helpful in increasing normality and heritability for right-skewed distributions such as somatic cell count, but for number of lactations, the transformation was not helpful because long survival times provide more information. For CLI and FLI, longer interval times provide information of poor fertility.

Censored records have to be included in the analysis

The exclusion of non-pregnant cows in the linear model analysis for CLI (L3) gave lower correlation between TBV_{CR} and PBV. Therefore, it is not recommendable to exclude censored records when using linear models. Furthermore, variance component estimates from L3 were lower, because part of the genetic variation was ignored. Donoghue *et al.* (2004) evaluated three methods for handling censored records for days to calving in beef fertility field data using a mixed model approach. They reported that when censored records were excluded from the analysis a lower estimate for additive variance was obtained. The authors highlighted the need to include records from non-calving cows in order to estimate differences in fertility for sires.

A smaller progeny group gave similar results

A large progeny group was used in the analyses of both traits. The progeny group size is similar to the Swedish breeding program, in which 120-150 daughters of young bulls are tested for functional traits. To study the effect of a smaller progeny group size, all models were analyzed using a progeny group size of 60 for FLI. However, the patterns in the correlations between TBV_{CR} and PBV were observed.

Calculating accuracy from estimated heritability works well if the model is appropriate

Table 1 shows accuracies of selection (r_{TI}) calculated from the estimated heritabilities and number of daughters for both traits. Weibull models S3_I (CLI), and S2 and S3_II (FLI) showed good agreement between the calculated accuracies and correlations between TBV_{CR} and PBV, which can be seen as the true

accuracy. This indicates that the heritability was estimated correctly for these models. However, this was not the case for the CLI Weibull model S2. If model S2 had been chosen as the "best" model based on the accuracy, it would have been the wrong choice because the heritability and thus r_{TI} was overestimated. The Cox model (S1) showed good agreement between the calculated accuracies and correlations between TBV_{CR} and PBV for CLI, but not for FLI, indicating that the heritability for FLI was not correctly estimated. For both CLI and FLI, heritabilities were not correctly estimated in model S4, in which all records were treated as uncensored.

For LM (L1 and L2), the accuracies and correlations between TBV_{CR} and PBV differed: accuracies were overestimated. Although the correlations between TBV_{CR} and PBV were higher for models L2 than L1, the log transformation of the data resulted in lower calculated accuracies. For models L1 and L2, heritabilities were therefore not correctly estimated.

Culling for reproduction gave bias

One hypothesis of the study was that when culling for reproduction is not properly handled using LM, breeding values of bulls with low genetic merit for daughter fertility are overestimated (bulls appear better than they really are). Figure 3 shows the prediction error, defined as the difference between PBV and TBV_{CR}, plotted against TBV_{CR}, for one replicate for CLI. PBV and TBV_{CR} were standardized by dividing each by its corresponding standard deviations. The best models are shown, model L2 and model S3_I (Weibull shifted).

Breeding values for the worst bulls for fertility (negative TBV_{CR}) were overestimated (positive prediction error) by LM. For SA the bias was small. Bias was also found for the best bulls for fertility; breeding values were underestimated. Similar results were also found for FLI.

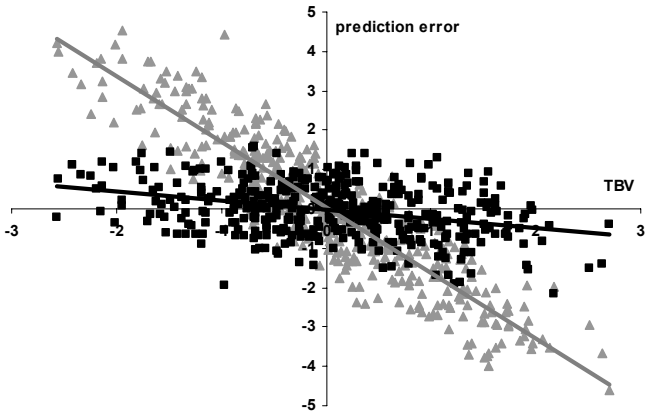

Fig. 3. Prediction error (PBV-TBV_{CR}) plotted against TBV for CLI for one replicate.
■ =Model S3_I; ▲= model L2; **----** = linear trend model S3_I; ---- = linear trend model L2.

Bias in breeding value prediction for bad bulls for fertility was expected in the LM due to culling for reproduction. SA predicted breeding values more accurately. Both results were in accordance with the hypothesis. However, there was an underestimation of PBV for good bulls for fertility, which was not expected. The reason for this is not clear.

FLI is a better measure than CLI for conception rate

FLI seems to be a slightly better trait to measure conception rate than CLI, based on the correlations between PBV and TBV_{CR}. In the simulation study, CLI depended also on the variation in the interval between calving and first ovulation. In field data, additional sources of variation could influence CLI, such as the interval between calving and first insemination, and management practices.

Sire vs animal model

A simple sire model was used to analyze both traits. Only information on daughters of sires (one batch of bulls) was available, and thus selection over time was not taken into account. Therefore it was not considered relevant to use an animal model; little gain is expected from an animal model for a lowly heritable trait with little selection pressure. Breeding values for cows were not interesting *per se*. However, technically it is possible to use an animal model with the Survival Kit if variance components are assumed known (Ducrocq, 2004).

Results from other studies

The application of SA to predict sire breeding values for fertility traits has been limited so far. Weigel (2004) studied two fertility traits in North American herds using survival models: days from calving until first positive pregnancy examination, with a Weibull model; and number of services to conception, with a discrete proportional hazards model. He reported a sire variance of 0.013 for the trait days from calving until first positive pregnancy examination, which is similar to the CLI results. In his study a threshold model (TM) was used to analyze the binary trait veterinarian-confirmed conception rate. Unfortunately no comparison was made to check which approach was the best to predict sire breeding values. However, Weigel stressed that SA is interesting due to its ability to provide a powerful, theoretically defensible analysis of interval traits, such as calving interval, days open, days to first service, calving until first positive pregnancy examination, which are subject to censoring.

Gonzáles-Recio *et al.* (2005) analyzed the number of inseminations to conception in Spanish Holsteins with a grouped model (discrete proportional hazards analysis). They reported a sire variance of 0.011 and a herd variance of 0.131 which is in the range of the FLI estimates (model S3_II). They also developed two other methods: an ordinal TM that accommodates censored records and a sequential TM which analyzes categorical traits that occur in a sequential order. From the cross validation method (taking random samples of herds from data) the proportional hazards and sequential TM provided more accurate

predictions for success at 1st insemination than the ordinal TM. However, the ordinal TM predicted probability of conception more accurately in subsequent inseminations. They stated that the results were surprising because both the sequential TM and the proportional hazards were expected to deal more properly with time-to-event traits, time-dependent covariates and censoring. They concluded that due to the categorical nature of the trait, the proportional hazards model might not be a proper specification as it assumes continuously recorded values that are grouped into categories.

What other models can be used?

Recently other methods have been proposed to analyze fertility traits. Averill *et al.* (2004) applied a longitudinal Bayesian TM of insemination events during the first 250 days after calving in first parity cows. This approach has some advantages: it allows for the inclusion of all breeding information within an opportunity period and can accommodate censoring. Furthermore it makes possible the joint analysis of male and female fertility and allows the use of an animal model.

Chang *et al.* (2006) used a bivariate censored threshold-linear model to estimate heritabilities and genetic correlations between number of services to conception and days open in first lactation Norwegian Red cows. The advantages of using this approach are that censored models take into account the uncertainty of pregnancy for culled cows, and that the bivariate analysis increases the accuracy of the predicted breeding values.

Multiple trait analysis

One potential disadvantage of using SA is that it has not been possible to analyze several traits together (*e.g.* production or other fertility traits) in a multiple-trait analysis to account for potential bias due to culling or selection over time. Recently, Damgaard (2005) has shown that it is indeed possible to analyze a survival trait together with a normally distributed continuous trait or a threshold trait using a Bayesian approach and applying Gibbs sampling. For large scale applications, approximations have been proposed (Ducrocq *et al.,* 2001; Tarrés *et al.*, 2006).

The importance of knowing pregnancy status

To take full advantage of survival analysis it is necessary to have information on conception so that censoring can be correctly specified. Insemination records collected by milk recording agencies are often incomplete and insemination data are not always properly recorded; indeed, sometimes only the successful insemination is recorded. Information on fertility, such as actual voluntary waiting period, service period, and pregnancy status, will hopefully be more accurately recorded in the future. Better quality data and the use of SA can be expected to give more than 10% greater genetic response than the use of LM. However, more research is needed using field data and incorporating other effects in the model.

Mastitis (Paper III)

The objective of the study was to compare the current method of genetic evaluation of clinical mastitis in Swedish dairy cattle with the SA approach. The higher heritabilitiy estimates for SA (Table 2) might partly be explained by the different opportunity periods used for the two traits, up to 700 d for TFM (SA) compared with 150 d for MAST (LM), and that TFM is more continuously distributed. The shorter opportunity period of 150 days was introduced in an effort to give all cows the same opportunity period as most of the culling (*e.g.*, for reproduction) occurs later in the lactation. With SA and the possibility to account for censoring, it is not necessary to define an opportunity period. However, it is debatable whether the longer opportunity period for SA has affected the results to a large extent since most of the first mastitis cases occurred within 150 d of lactation (70, 66 and 68% for all cases for 1st, 2nd and 3rd lactation, respectively).

The higher accuracies for SA can be translated into a higher genetic response, approximately, 3% for first parity and 25% for later parities. The large differences in accuracy between the first and later lactations could be related to culling due to high SCC. The incidence of mastitis and SCC increase with increasing parity number. Cows culled for high SCC are treated as censored in SA, whereas they are treated as healthy with LM.

Results from other studies using SA

Some studies using SA to analyze clinical mastitis have been done. Saebø *et al.* (2002) analyzed time to first mastitis treatment on a small dataset of the first 5 lactations in Norwegian cattle, with both a stochastic process model and a semi-parametric proportional hazards model. They concluded that the stochastic process model seemed to be better for later stages of lactation, whereas the semi-parametric model was better around calving, and they were able to identify sires with daughters showing highest resistance to mastitis. They did not compare these models with linear models. Saebø & Frigessi (2004) used a proportional hazards model for modeling time to first mastitis. They found considerable differences between sires with regard to their daughters' disease resistance. The authors did not address heritability and sire variance component estimation. Saebø *et al*. (2005) analyzed time to first treatment of clinical mastitis using a competing process model as an alternative to the proportional hazards models. They found that the properties of the estimated latent process harmonized with prior biological knowledge. However, they pointed out that this model is computationally demanding.

There are some drawbacks of applying SA for the analysis of clinical mastitis. First, it is not easily possible to apply it in a multiple-trait analysis, and a genetic evaluation together with SCC would be desirable. However, as was mentioned for the fertility traits, some alternatives have been proposed. Second, only the first case of mastitis is considered in the definition. This might be of less importance as most treated cows have only one case of mastitis.

Other methods proposed to study mastitis

TMs have been proposed to analyze clinical mastitis although they are not used for national routine genetic evaluations. TM accounts for the binary nature of the trait and can be more advantageous for variance component and breeding value estimation (Gianola and Foulley, 1983). Heringstad *et al.* (2001) used a cross-sectional threshold model, which, like the cross-sectional LM, only considers the first case of mastitis.

Bayesian longitudinal (Heringstad *et al.,* 2003; Chang *et al.,* 2004) and multivariate TM (Heringstad *et al*., 2004) have been proposed. The advantage of these approaches is that they take into account multiple treatments of clinical mastitis and time aspects. Moreover, records in progress and incomplete records due to culling can be accommodated in the analysis. Heringstad *et al. (*2003) stated that the longitudinal approach may not be effective in capturing differential gene expression in different parts of the lactation because the genetic or sire variance-covariance structure is static (the Legendre function does not have a genetic component). Chang *et al.* (2004) reported that the time covariate strongly affected results, and genetic parameters should be therefore be interpreted with caution. Furthermore, these approaches need more records per animal and more parameters, resulting in the methods being computationally demanding.

A drawback in animal breeding is the lack of suitable statistical methods to compare results from LM, SA and other approaches when using field data. The advantage of a simulation study is that true breeding values are known, which is not the case in real life. The comparison using correlations between predicted and true breeding values is the best method to compare different approaches.

Differences in the application of SA to fertility and mastitis

The application of SA in the genetic evaluation of female fertility (CLI and FLI) and clinical mastitis (TFM) gave different results. SA worked quite well to analyze fertility traits. However, for the analysis of clinical mastitis the differences between the two approaches were not so large, especially for first parity, given the theoretical advantages of SA in handling censoring.

A simulation study, in which true breeding values are known and can be correlated with breeding values predicted by different approaches, can provide additional information about the suitability of different approaches to predict breeding values. Carlén *et al.* (2006) investigated by simulation whether SA (using TFM) resulted in a more accurate genetic evaluation for clinical mastitis than LM and TM (using MAST). They found that correlations between true breeding values for mastitis liability and breeding values predicted by LM, SA and TM were almost the same, and concluded that little would be gained by replacing LM with the SA or TM, although both SA and TM have theoretical advantages. These results were somewhat unexpected. A large progeny group was used in the simulation; therefore with a lot of information all models worked quite well.

Another explanation could be that in the simulation culling was not related to mastitis liability which it might be in real life.

Another possible reason for the difference of performance of SA for fertility and mastitis could be related to the distribution of the traits. For CLI, observations start at day 56 and onwards, and for FLI 50% of the cows conceived at the first insemination and the other cows potentially conceived at intervals of 21 days on average (Figure 2 in Paper I and Figure 1 in Paper II, respectively). For clinical mastitis (Figure 4), most of the first cases of mastitis treatment occurred in a short period around calving. In theory, TFM is a better trait to describe clinical mastitis because the timing of mastitis within the lactation is considered and TFM is more continuously distributed. However, most of the mastitis cases occur at the same time, making TFM behaves almost as a binary trait. In other words, cows either do or do not get mastitis during that early period. The almost binary nature of TFM could explain why SA has no extra advantage over LM or TM.

The proportion of censoring was different for fertility and mastitis and this could have had an effect as well. CLI and FLI had a high proportion of uncensored records (approximately 85%), which correspond to pregnant cows. These records have complete information and therefore give rise to more reliable estimates. The inverse situation was found for the incidence of first treatment of clinical mastitis. Approximately 15% of the records were uncensored, which correspond to cows that have been treated or have been culled due to mastitis. This resulted in a high proportion of censoring (incomplete information), which might have lowered the reliability of estimates.

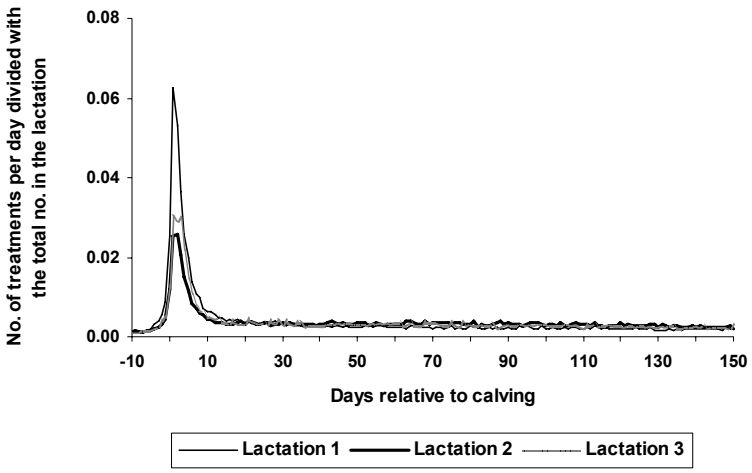

Fig. 4. Relative frequency of mastitis in 150 days of the three first lactations (taken from Carlén, 2003).

Effects of pregnancy status and mastitis on culling (Paper IV)

The interesting and novel result of this study was the interaction between the effects of pregnancy status and mastitis, which has not been reported before. The results were quite consistent with other studies in which pregnancy status and mastitis were considered as main effects in the models (Beaudeau *et al.,* 1995; Gröhn *et al.,* 1998; Rajala-Schultz & Gröhn, 1999b; Schneider *et al.,* 2005).

The inclusion of an interaction term allowed us to identify and quantify the risk of culling for healthy and sick cows, treated at different stages of the lactation, accounting for the knowledge of pregnancy status at different stages of the lactation. Mastitis affected culling decisions throughout the lactation, but its effect depended on pregnancy status and the stage during which the cow was treated. The risk of culling was low for pregnant cows in mid lactation, whether or not they had been treated for mastitis. Among the non-pregnant cows, the risk of culling became greater the later in the lactation the cow had become pregnant. Different patterns were observed for the risk of culling between open and pregnant cows. For both groups, the risk of being culled was higher for cows treated for mastitis than for untreated cows. For the open group, the risk was higher for cows treated in earlier stages, whereas for the pregnant group, the risk was more similar across different stages of lactation.

For practical reasons, a relatively old data set was used for this study. The original research project focused on genetic studies for longevity (Roxström, 2001, PhD thesis) and reasons for culling in Swedish dairy cattle. The data included information until 1996, to avoid problems due to the introduction of the quota milk system in Sweden. However, the quota would not have affected the estimates because Sweden never reached the quota assigned by the EU. It would be interesting to apply the studied models to more recent data, to estimate parameters and to compare culling policies for different periods of time.

Survival analysis was found to be a useful method to analyze the effects of pregnancy status and clinical mastitis as risk factors for culling by treating them as time-dependent covariates in the model. The possibility of accounting for the timing of both effects during different periods of the lactation might have lead to more accurate estimates. Indeed it would not have been possible to develop such an analysis with another method.

The knowledge of how the timing of treatment of clinical mastitis and pregnancy status affects culling risk during the lifetime of a cow can be used to identify periods of great risk. Technicians and researchers can thereby focus on finding better herd management practices for these periods. If successful, the incidence of impaired fertility and mastitis would be reduced.

The estimates from SA can be used in economic simulation studies to calculate costs associated to the effects caused by the interaction of pregnancy status and mastitis, which might be different when these two effects are evaluated separately. As was shown in the study, different culling patterns were observed as a result of the interaction.

Final considerations

This thesis has shown the feasibility of applying SA to predict breeding values for female fertility and clinical mastitis traits. The application of methods that can handle problems related to the distribution and characteristics of these traits is recommendable. More accurate breeding values were predicted and thus higher genetic gain can be expected if SA is used.

SA was also found to be a suitable method to study the effects of pregnancy status and mastitis, and their interaction on longevity. The inclusion of these effects as time-dependent covariates made it possible to account for the timing of their incidence. Thus, more accurate solutions were obtained and more refined knowledge of their effect on culling was gained.

Conclusions

Survival analysis was a better approach than LM to predict sire breeding values for conception rate when using observations on CLI and FLI, and for clinical mastitis when using observations on TFM:

- Correlations between TBV_{CR} and sire breeding values for CLI and FLI predicted with survival analysis were higher than the corresponding correlations from linear models. If selection were carried out on these PBVs, it would translate into higher genetic progress, 8 to 12% for FLI and CLI, respectively.

- Breeding values predicted by Weibull S3_I (CLI) and the grouped data model S3_II (FLI) had the highest correlations with TBV_{CR} and the calculated accuracies showed good agreement.

- Accuracies in selection were higher for SA, using TFM, compared with LM, using MAST, for all parities. The higher accuracy could be translated into a higher genetic progress, 3 and 25% in first and later parity, respectively.

Pregnancy status and clinical mastitis were important risk factors for longevity in Swedish dairy cattle. An interaction between these two effects was found and its effect was quantified:

- The farmer's knowledge of pregnancy status, in particular, strongly affected longevity: the risk of culling in pregnant cows fell sharply, whether the cow was healthy or not. However, the later in lactation a cow became pregnant, the greater was her risk of being culled.

- Mastitis affected culling throughout the lactation. For open cows the effect of mastitis was more marked in early lactation, whereas for pregnant cows the risk of culling was similar across the different stages of lactation.

Future research

More research is suggested to further evaluate the suitability of SA for genetic evaluation for fertility and clinical mastitis traits. Some suggestions are:

Fertility

- To consider more variation for some factors, such as voluntary waiting period and heat detection rate within and across herds in the simulation

study. For instance, it is reasonable to expect the voluntary waiting period to be affected by the cow's milk production level.
- To include relationships among animals and selection over time (generations), and apply an animal model.
- To include other factors in the models, such as age of calving, season of calving, and lactation stage.
- To explain why the Weibull model seems to work so well for FLI, even when the assumption of the Weibull distribution does not hold very well.
- To apply SA in field data.
- To study the impact of censoring when new daughters of young bulls are evaluated.

Clinical mastitis

- To study different opportunity periods for both clinical mastitis traits using field data.
- To include a time-dependent lactation stage effect in the model, to account for the much higher frequency of mastitis around calving.
- To consider a lactation basis model, in which each lactation is treated separately.

For both traits

- To develop statistical methods to compare different approaches when using field data.
- To evaluate the cost, benefits, and resources needed to implement sophisticated methods in large scale applications.
- Application of a multiple trait approach, analysis of survival traits together with normally distributed traits.

Regarding risk factors for longevity, other aspects that could be interesting to study are:

- What is the risk of culling in herds with high or low levels of somatic cells counts or high and low incidence of clinical mastitis?
- How do multiple treatments of clinical mastitis affect culling?
- Do culling policies in high and low producing herds differ?
- How does the incidence of other disorders (*e.g.,* metabolic diseases) affect longevity?
- Can the study of specific culling reasons (competing risk analysis) give more insight into the culling process?

References

Andersson, L., Gustafsson, A.H., Emanuelson, U. 1991. Effect of hyperketonemia and feeding on fertility in dairy cows. *Theriogenology* 36 (4), 521-536.

Averill, T.A., Rekaya, R. & Weigel, K. 2004. Genetic analysis of male and female fertility using longitudinal binary data. *Journal of Dairy Science* 87, 3947-3952.

Beaudeau, F., Ducrocq, V., Fourichon, C. and Seegers, H. 1995. Effect of disease on length of productive life of French Holstein dairy cows assessed by survival analysis. *Journal of Dairy Science* 78, 103-117.

Beaudeau, F., Seegers, H., Ducrocq, V., Fourichon, C., Bareille, N. 2000. Effect of health disorders on culling in dairy cows: a review and critical discussion. *Annales de Zootechnie* 49, 293-311.

Carlén, E. 2003. Genetic parameters of mastitis traits and production in the first three lactations of Swedish Holsteins cows. University of Agricultural Sciences, Uppsala, Sweden. *Examensarbete 244.*

Carlén, E., Emanuelson, U. & Strandberg, E. 2006. Genetic evaluation of mastitis in dairy cattle using linear models, threshold models, and survival analysis: a simulation study. *Journal of Dairy Science.* Submitted.

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.

Cox, D.R. 1972. Regression models and life tables (with discussion). *Journal of the Royal Statistical Society. Series B, Statistical Methodology.* 34, 187.

Chang, Y.M., Gianola, D., Heringstad, B. and Klemetsdal, G. 2004. Longitudinal analysis of clinical mastitis at different stages of lactation in Norwegian cattle. *Livestock Production Science* 88, 251-261.

Chang, Y.M., Andersen-Ranberg, I.M., Heringstad, B., Gianola, D. and Klemetsdal, G. 2006. Bivariate analysis of number of services to conception and days open in Norwegian Red using a censored threshold-linear model. *Journal of Dairy Science* 89, 772-778.

Donoghue, K.A., Rekaya, R. & Bertrand, J.K. 2004. Comparison of methods for handling censored records in beef fertility data: field data. *Journal of Animal Science* 82, 357-361.

Damgaard, L.H. 2005. Quantitative genetic analysis of survival, linear Gaussian and ordered categorical traits. *PhD thesis.* The Royal Vet. and Agric. Univ., Denmark.

Ducrocq, V. & Sölkner, J. 1998. The Survival Kit - a Fortran package for the analysis of survival data. *Proceedings 6th World Congress on Genetics Applied to Livestock Production, Armidale, Australia.* 27, 447-448.

Ducrocq, V., Boichard, D., Barbat, A., Larroque, H. 2001. Implementation of an approximate multitrait BLUP evaluation to combine production traits and functional traits into a total merit index. *Proceedings 51st Annual Meeting of the European Association for Animal Production. Book of Abstracts,* pp. 2.

Ducrocq, V. 2004. An improved model for the French genetic evaluation of dairy bulls on length of productive life of their daughters. *Proceedings 55st Annual Meeting of the European Association for Animal Production. Book of Abstracts,* pp. 91.

Gianola, D., Foulley, J.L. 1983. Sire evaluation for ordered categorical data with a threshold model. *Genetics Selection Evolution* 15, 201-224.

González-Recio, O., Chang, Y.M., Gianola, D. and Weigel, K.A. 2005. Number of inseminations to conception in Holstein cows using censored records and time-dependent covariates. *Journal of Dairy Science* 88, 3655-3662.

Gröhn, Y.T., Ducrocq, V. & Hertl, J.A. 1997. Modeling the effect of a disease on culling: an illustration of the use of time-dependent covariates for survival analysis. *Journal of Dairy Science 80,* 1755-1766.

Gröhn, Y.T., Eicker, S.W., Ducrocq, V. and Hertl, J.A. 1998. Effect of diseases on the culling of Holstein dairy cows in New York state. *Journal of Dairy Science* 81, 966-978.

Harman, J.L., Gröhn, Y.T., Erb, H.N., Casella, G. 1996a. Event-time analysis of the effect of 60-day milk production on the parturition-to-conception interval in dairy cows. *American Journal of Veterinary Research* 57 (5), 634-639.

Harman, J.L., Gröhn, Y.T., Erb, H.N., Casella, G. 1996b. Event-time analysis of the effect of season of parturition, parity, and concurrent disease on parturition-to-conception interval in dairy cows. *American Journal of Veterinary Research* 57 (5), 640-645.

Heringstad, B., Rekaya, R., Gianola, D., Klemetsdal, G. and Weigel, K.A. 2001. Bayesian analysis of liability of clinical mastitis in Norwegian cattle with a threshold model: effects of data sampling method and model specification. *Journal of Dairy Science* 84, 2337-2346.

Heringstad, B., Chang, Y.M., Gianola, D. and Klemetsdal, G. 2003. Genetic analysis of longitudinal trajectory of clinical mastitis in first-lactation Norwegian cattle. *Journal of Dairy Science* 86, 2676-2683.

Heringstad, B., Chang, Y.M., Gianola, D. and Klemetsdal, G. 2004. Multivariate threshold model analysis of clinical mastitis in multiparous Norwegian dairy cattle. *Journal of Dairy Science* 87, 3038-3046.

Interbull. 2006. Description of National Genetic Evaluation Systems for dairy cattle traits as
applied in different Interbull member countries. Interbull Centre, Sweden. http://www-interbull.slu.se/national_ges_info2/framesida-ges.htm (Accessed 24-Mar-2006).

Jensen, J. & Madsen, P. 1994. DMU: A package for the analysis of multivariate mixed models. *Proceedings 5th World Congress on Genetics Applied to Livestock Production, Guelph, Canada. Computing Strategies and Software* 22, 45-46.

Jensen, J., Mäntysaari, E.A., Madsen, P. and Thompson, R. 1997. Residual maximum likelihood estimation of (co)variance components in multivariate mixed linear models using average information. *Journal of the Indian Society of Agricultural Statistics* 49, 215-236.

Jorjani, H. 2005. Preliminary report of Interbull pilot study for female fertility traits in Holstein populations. *Proceedings of the Interbull meeting Uppsala, Sweden. Interbull Bulletin* 33, 34 – 44.

Lee, E.T. 1992. *Statistical methods for survival data analysis.* 2nd ed., John Wiley & Sons, Inc., New York, NY. 482 pp.

Lee, L.A., Ferguson, J.D. & Galligan, D.T. 1989. Effect of disease on days open assessed by survival analysis. *Journal of Dairy Science* 72, 1020-1026.

Mark, T., Nielsen, U.S., Pösö, J., Gundel, M., Svendsen, M. 2001. Genetic relationships among functional traits in the Nordic Holstein populations. *Proceedings of the 2001 Interbull Meeting in Budapest, Hungary. Interbull Bulletin* 27, 64-67.

Neerhof, H.J., Madsen, P., Ducrocq, V.P., Vollema, A.R., Jensen, J. and Korsgaard, I.R. 2000. Relationships between mastitis and functional longevity in Danish Black and White dairy cattle estimated using survival analysis. *Journal of Dairy Science* 83, 1064-1071.

Prentice, R. & Goeckler, L. 1978. Regression analysis of grouped survival data with application to breast cancer data. *Biometrics* 34, 57-67.

Rajala-Schultz, P.J. & Gröhn, Y.T. 1999a. Culling of dairy cows. Part I. Effects of diseases on culling in Finnish Ayrshire cows. *Preventive Veterinary Medicine* 41, 195-208.

Rajala-Schultz, P.J. & Gröhn, Y.T. 1999b. Culling of dairy cows. Part II. Effects of diseases and reproductive performance on culling in Finnish Ayrshire cows. *Preventive Veterinary Medicine* 41, 279-294.

Roxström, A. 2001. Genetics aspects of fertility and longevity in dairy cattle. *PhD thesis*. University of Agricultural Sciences. Uppsala, Sweden.

SAS 1999. *SAS Release V8.1.* 1999-2000. SAS Inst., Inc., Cary, NC.

Saebø, S., Almoy, T., Aastveit, A.H., Heringstad, B. and Klemetsdal, G. 2002. Modelling time to first treatment of clinical mastitis as first passage times of stochastic processes. *Proceedings 7th World Congress on Genetics Applied to Livestock Production* Montpellier, France. *CD-ROM communication* Nº 16-16.

Sæbø, S. & Frigessi, A. 2004. A genetic and spatial Bayesian analysis of mastitis resistance. *Genetics Selection Evolution* 36, 527-542.

Sæbø, S., Almøy, T., Heringstad, B., Klemetsdal, G. and Aastveit, A.H. 2005. Genetic evaluation of mastitis resistance using a first-passage time model for Weiner processes for analysis of time to first treatment. *Journal of Dairy Science* 88, 834-841.

Schneider M. del P., Strandberg, E. & Roth, A. 2005. Effect of mastitis and fertility on culling in Swedish dairy cattle. *Proceedings 56th Annual Meeting of the European Association for Animal Production*. *Book of Abstracts,* pp. 174.

Strandberg, E. & Shook, G.E. 1989. Genetic and economic responses to breeding programs that consider mastitis. *Journal of Dairy Science* 72, 2136-2142.

Suriyasathaporn, W., Nielsen, M., Dieleman, S.J., Brand, A., Noordhuizen-Stassen, E.N.,

Schukken, Y.H. 1998. A Cox proportional-hazards model with time-dependent covariates to evaluate the relationship between body-condition score and the risks of first insemination and pregnancy in a high-producing dairy herd. *Preventive Veterinary Medicine* 37, 159-172.

Swedish Dairy Association. 2002. *Husdjursstatistik (Cattle Statistics) 2002*. Svensk Mjölk. SE-631 84 Eskilstuna, Sweden.

Tarrés, J., Piedrafita, J. & Ducrocq, V. 2006. Validation of an approximate approach to compute genetic correlations between longevity and linear traits. *Genetics Selection Evolution* 38, 65-83.

VanRaden, P.M. 2003. Longevity and fertility traits definitions compared in theory and simulation. *Proceedings of the Interbull Technical Workshop Beltsville, MD, USA*. *Interbull Bulletin* 30, 43-46.

VanRaden, P.M., Sanders, A.H., Tooker, M.E., Miller, R.H., Norman, H.D., Kuhn, M.T. and Wiggans, G.R. 2004. Development of a national genetic evaluation for cow fertility. *Journal of Dairy Science* 87, 2285-2292.

Weigel, K.A. 2004. Improving the reproductive efficiency of dairy cattle through genetic selection. *Journal of Dairy Science* 87 (E. Suppl.), E86-E92.

Yazdi, M.H., Visscher, P.M., Ducrocq, V. and Thompson, R. 2002. Heritability, reliability of genetic evaluations and response to selection in proportional hazard models. *Journal of Dairy Science* 85, 1563-1577.

Acknowledgements

I would like to express my sincere gratitude to:

Erling Strandberg, for being an excellent supervisor, for your valuable guidance, discussions and constant support.

My co-supervisors: Anki Roth, Ulf Emanuelson and Katja Grandinson, for your help and contributions.

Vincent Ducrocq, for your help, ideas and for being always available to answer all my questions.

Swedish Farmers Foundation for Agricultural Research and the SLU research theme "Animal welfare for quality in food production" for financing part of the project.

Knut och Alice Wallenbergs Stiftelse for the financial support to participate in the 55th EAAP meeting in Slovenia, 2004.

Swedish Dairy Association for providing the data used in the studies.

All the staff and PhD students (past and present) of the HGEN department, for being friendly and always ready to help me.

All my friends over these years in Uppsala and my multi-across-country friends.

My parents, sisters (Matilde, thanks for the beautiful painting), brothers and nephews, for their support.