MODELLING TIME TO FIRST TREATMENT OF CLINICAL MASTITIS AS FIRST PASSAGE TIMES OF STOCHASTIC PROCESSES

S. Saebo[1], T. Almoy[1], A.H. Aastveit[1], B. Heringstad[2] and G. Klemetsdal[2]

[1]Department of Mathematical Sciences, Agricultural University of Norway, Ås, Norway
[2]Department of Animal Science, Agricultural University of Norway, Ås, Norway

INTRODUCTION

Mastitis is the most frequent and costly disease affecting dairy cattle. The disease shows different infection patterns and is associated with many risk factors. Due to this complexity it may be appropriate to approximate the unobserved development of the disease by a stochastic process. A stochastic process operate in some state space and move from one state to another with a certain probability. The state space can have an absorbing state which terminates the process if reached, and onset of a disease may be viewed as an absorbing state. The time from the process is initiated until it is absorbed (the first passage time) is a stochastic variable following some probability distribution.

The risk of mastitis is roughly constant throughout most of the lactation period, except the days around calving when the risk is highly increased (Heringstad *et al.,* 1999). It is believed that this increase may be due to immunological changes (Kehrli *et al.,* 1989). Since there are several risk factors involved, it may be appropriate to assume the presence of competing latent processes in a competing risks model. The difference from usual competing risks models is that we do not know which risk factors that cause the onset of the disease. For methods dealing with latent competing risks in the case of proportional hazards models we refer to Gelfand *et al.,* (2000).

The objective of this study was to analyse time to first mastitis with a stochastic process model and to compare this model with a semi-parametric proportional hazards model which have become popular in the area of animal breeding due to the work of Ducrocq and Casella (1996).

MATERIAL AND METHODS

Data. The data was extracted from the data set analysed by Heringstad *et al.,* (1999). Records from 3 454 daughters of 201 Norwegian Cattle (NRF) sires were used. The cows had first calving in 1990, 1991 and 1992 and were from 260 herds. Time to first veterinary treatment of clinical mastitis was calculated within each lactation, including 31 days prior to calving, for each cow, up to 5 lactations. In the resulting data set containing 6 618 records, the number of observations on each cow thus varied from 1 to 5. Records for cows entering next lactation before first treatment of mastitis was treated as censored. A total of 302 male ancestors, including the 201 sires with daughter records, were represented in the pedigree file.

Statistical models. Figure 1a) shows two simulated examples of wiener processes operating in the continuous state space $[0, \infty)$, with 0 as an absorbing state. One process was initiated in state $c_1 = 5$, whereas the second was initiated in state $c_2 = 10$. The second process has drift $\mu >$

0 towards the absorbing state 0 and was absorbed at time $t \approx 255$. The first process has negative drift, away from absorption. The figure also indicates that the initiation of a process was allowed to be delayed, as in the case of the second process which was delayed by $\varphi = 100$. The probability distribution $f_j(t_j|\theta_j)$ for the first passage time t_j of a process j depends on a vector of parameters θ_j and the starting conditions of the process. Two process scenarios were used ; I) Processes initiated in a single state c with probability 1 and drift μ, leading to Inverse Gaussian distributed first passage times (Chhikara and Folks, 1989), and II) Processes initiated according to a probability distribution over the state space. Under certain conditions it can be shown that the resulting first passage time distribution is an exponential distribution with expectation μ^{-1} ($\mu > 0$) (Aalen and Gjessing, 2001). Under both scenarios the process initiation may be delayed by φ time units.

The hazard function expresses the instantaneous risk of an event to happen at some time t given that it has not occurred prior to t. The process specific hazard function is given by $h_j(t_j|\theta_j) = f_j(t_j|\theta_j)/S_j(t_j|\theta_j)$, where the survivor function $S_j(t_j|\theta_j)$ expresses the probability that the process has not been absorbed prior to t_j. On the assumption of k latent processes, corresponding to k risk factors, an overall first passage time is defined by $t = min(t_1,...,t_k)$. By assuming independent processes it is straight forward to show that the overall survivor function is the product of the process specific survivor functions, and that the overall hazard function is the sum of the process specific hazard functions. The likelihood function of the data will be specified through these.

The drift parameters of the processes were expressed as functions of the covariates, allowing for herd and sire dependent drift. The model adopted for the drift of the j'th process for record i is :

$$\mu_{ij} = g_j(\xi_j + s_{j,sire(i)} + h_{j,herd(i)} + \alpha_j * age_i + \beta_{j,year(i)} + \gamma_{j,month(i)} + \lambda_{j,lact(i)})$$

where $s_{j,sire(i)}$ is the effect of the sire of the cow with record i, $h_{j,herd(i)}$ is effect of herd, α is effect of age of cow at first calving (in days), $\beta_{j,year(i)}$ is effect of year of calving (3 classes), $\gamma_{j,month(i)}$ is effect of month of calving (12 classes) and $\lambda_{j,lact(i)}$ is effect of lactation number (5 classes). The vector of sire effects, $\boldsymbol{s}_j$ was assumed MVN($\boldsymbol{0}, \sigma_s^2 \boldsymbol{A}$), where $\boldsymbol{A}$ is the relationship matrix of the sires. The vector of herd effects, $\boldsymbol{h}_j$ was assumed MVN($\boldsymbol{0}, \sigma_h^2 \boldsymbol{I}$), where $\boldsymbol{I}$ is the identity matrix. Generally, the number of latent processes may be estimated by fitting models for various values of k and selecting the k maximizing some model selection criterion. In this paper, however, $k = 2$ was chosen assuming that process 1, for mastitis due to changes in animal physiology just prior to calving, is initiated according to scenario I) above, and that process 2 is initiated according to scenario II), resulting in constant hazard in the remaining lactation. For model comparison a semi-parametric proportional hazards model with a piecewise constant baseline hazard multiplied with the same covariates in a loglinear expression was fitted. The assumed distributions of the sire and the herd effects were as for the stochastic process model.

Estimation. A Bayesian approach was used for inferences. The model parameters were estimated by means of a Metropolis-Hastings algorithm (Metropolis *et al.,* 1953). For record i the observed data is denoted by (y_i, δ_i) where y_i is either a failure time with corresponding $\delta_i = 1$

or a censoring time with $\delta_i = 0$. For the general *k*-process model the likelihood function can be shown to be :

$$L(\theta \mid y) = \prod_{i=1}^{n} \left(\left[\sum_{j=1}^{k} h_j(y_i \mid \theta_j) I(y_i > \varphi_j) \right]^{\delta_i} \left[\prod_{j=1}^{k} S_j(y_i \mid \theta_j)^{I(y_i > \varphi_j)} \right] \right)$$

where $I(y_i > \varphi_j)$ is an indicator of y_i being larger than the delay time of the *j*'th latent process. For the proportional hazards model the likelihood was constructed according to Korsgaard *et al.,* (1998) and references therein. In both models gamma priors were assumed for the inverse of the variance components of the sire and herd effects, whereas normal or log-normal priors were assigned for the other parameters. The conditional predictive ordinate (CPO) (e.g. Ibrahim *et al.,* 2001) was computed for each observation and used for model comparison.

RESULTS AND DISCUSSION

Both models capture well the peaked hazard around calving. The model with the larger mean of log(CPO)'s predicts new data better (Ibrahim *et al.,* 2001) and the values obtained for the stochastic process model and the semi-parametric model were -1.733 and -1.753 respectively. Permutation tests showed, however, no overall difference in CPO-values. Figure 1b) shows the log of the CPO-ratios for each observation versus time, and values above zero support the stochastic process model. This model seems to be better in later stages of lactation (from day ≈ 100), while the semi-parametric model is superior at day 28 and 29, that is, 2 - 3 days before calving, in addition to a short period after calving. This indicates that the assumption of a constant hazard prior to calving for the stochastic process model probably is not appropriate.

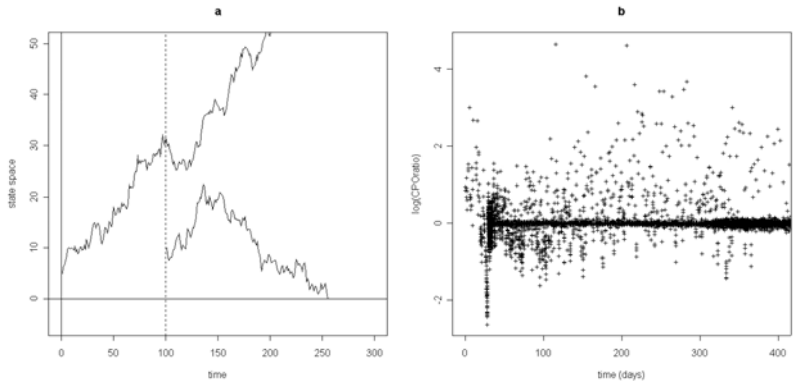

Figure 1. (a) Example of two simulated wiener processes. The first passage time for each process is when the process hits the state zero. (b) The logarithm of the conditional predictive ordinate (CPO)-ratios for the stochastic process model and the semi-parametric model for each observation versus time. Values above zero support the stochastic process model

It is reasonable to believe that the influence of covariates on risk factors varies from factor to factor which in turn leads to an overall non-proportional hazard. The stochastic process model allows for non-proportional hazards which is probably the reason why it fits data better in later lactation than the semi-parametric model. For instance, daughters of some sires showed high susceptibility to mastitis around calving, whereas daughters of other sires seemed more resistant in this period, but more susceptible in later lactation.

Mastitis shows a very complex hazard pattern throughout the lactation, and it may be difficult to fit parametric hazard functions when the hazard is irregular and non-monotone. The semi-parametric model handles this elegantly by assuming an arbitrary baseline hazard. However, the strong assumption of proportional hazards may be inappropriate. Complex and non-proportional hazards may indicate multiple hidden risk factors. By the methods used in this paper we were able to fit fully parametric models to the complex hazard pattern of mastitis by assuming latent stochastic processes approximating the unobserved development of the disease. Both genetic and environmental factors influence the disease directly, and this is mimicked by allowing covariates to act directly on the drift parameters of the processes. In this manner we were able to identify sires with daughters showing highest resistance to mastitis.

ACKNOWLEDGMENTS

The Norwegian Dairy Herd Recording (Husdyrkontrollen) is acknowledged for providing data. This work is financed by the Norwegian Research Council through the "Healthy Cow" project.

REFERENCES

Aalen, O.O. and Gjessing, H.K. (2001) *Stat. Sci.* **16** : 1-22.

Chhikara, R.S. and Folks, J.L. (1989) "The Inverse Gaussian Distribution : Theory, Methodology and Applications". Marcel Dekker, New York, USA.

Ducrocq, V. and Casella, G. (1996) *Genet. Sel. Evol.* **28** : 505-529.

Gelfand, A.E., Gosh, S.K., Christiansen, C., Soumerai, S.B. and McLaughlin, T.J. (2000) *Appl. Statist.* **49** : 385-397.

Heringstad, B., Klemetsdal, G. and Ruane, J. (1999) *J. Dairy Sci.* **82** : 1325-1330.

Ibrahim, J.G., Chen, M.-H. and Sinha, D. (2001) "Bayesian survival analysis", Springer-Verlag, New York, USA.

Kehrli, M.E., Nonnecke, B.J. and Roth, J.A. (1989) *Am. J. Vet. Res.* **50** : 207-214.

Korsgaard, I.R., Madsen, P. and Jensen, J. (1998) *Genet. Sel. Evol.* **30** : 241-256.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H. and Teller, E. (1953) *J. Chemical Physics* **21** : 1087-1091.