55th Annual Meeting of the European Association for Animal Production, Bled, Slovenia, September 5th – 9th, 2004, Commission on Animal Genetics, Session G2.8

Genetic Evaluation for Mastitis using Survival Analysis

E. Carlén, M. del P. Schneider, and E. Strandberg

Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, P.O. Box 7023, SE-750 07 Uppsala, Sweden
Corresponding author, e-mail: Emma.Carlen@hgen.slu.se

Abbreviation key: LM = linear models, **MAST** = mastitis (0/1), **SA** = survival analysis, **TFM** = time to first mastitis or censoring.

INTRODUCTION

Mastitis is one of the most common and costly diseases in dairy cattle, and due to its unfavorable genetic correlation with milk production (e.g. Heringstad et al., 2000; Hansen et al., 2002; Carlén et al., 2004) it is of great importance to include clinical mastitis or other correlated traits in the breeding goal. Currently, the most common method for genetic evaluation of clinical mastitis is linear model (**LM**) methodology, where mastitis is defined as a binary trait within a defined period of the lactation. The heritability of clinical mastitis analyzed with LM is low (e.g. Pösö and Mäntysaari, 1996; Rupp and Boichard, 1999; Carlén et al., 2004). This can mainly be explained by large environmental effects but also partly by the all-or-none character of the trait.

One of the drawbacks with the traditional LM for analyzing clinical mastitis is the low observed variation among cows connected to the all-or-none character of the trait. With this method there is no difference between cows getting mastitis in the beginning or close to the end of the defined time period, or between cows without mastitis cases. Furthermore, cows culled before they got a chance to express mastitis are either treated as missing observations or as healthy cows, which cannot be separated from cows that did not contract mastitis during the whole period. Consequently, the amount of information used is reduced when mastitis is analyzed as an all-or-none trait with LM.

Survival analysis (**SA**), also known as failure-time or event-time analysis, is a statistical method for studying the occurrence and timing of specific events, where the analyzed response time equals the time elapsed from a starting point until the occurrence of the event of interest (Ducrocq, 1987). SA has been frequently and successfully used for genetic analyses of dairy cattle longevity traits (e.g. Smith and Quaas, 1984; Ducrocq, 1994), whereas genetic studies of the time to first outbreak of disease by the use of SA, or similar methodology, have so far been limited. Saebo et al. (2002) analyzed time to first mastitis treatment on a relatively small dataset of Norwegian cattle, with both a stochastic process model and a semi-parametric proportional hazard model, and Hirst et al. (2002) studied time to occurrence of lameness in dairy cattle.

Considering the advantages of SA for traits with a longitudinal character, improved efficiency is also expected for mastitis data when time to first case of clinical mastitis (**TFM**) is analyzed. The observed variation among cows, and among sires, increases, since all cows with a mastitis case would not automatically get the same value but a value affected by the timing of the case. Cows without cases and cows culled (for reasons other than but possibly correlated to mastitis) before they got a chance to express mastitis, are treated as censored records. The latter reduces the potential bias occurring when treating these cows as healthy with traditional methodology.

The objective of this study was to investigate whether TFM could be successfully analyzed using survival analysis and to compare the precision of predicted breeding values from that analysis to the precision from a traditional linear model.

MATERIALS AND METHODS

Data

Data were extracted from the Swedish milk recording scheme, and were edited to include records from the first 3 lactations of Swedish Holstein cows having their first calving between 1995 and 2000. Further it was restricted to include only: 1) cows calving between 20 and 38, 32 and 52, and 43 and 66 months of age at first, second, and third calving, respectively, 2) cows belonging to a herd-year class with at least 2 observations, 3) cows from sires with at least 50 daughters in the data before editing. The structure of the analyzed data for lactation 1 to 3 is shown in Table 1. The sire pedigree file had 1139 bulls, including the sires with daughter records.

For LM, mastitis was defined, in the same way as in the Swedish national genetic evaluation, as a veterinary treated clinical mastitis from 10 d before to 150 d after calving, or culling for mastitis within that period. The restricted time period was introduced to reduce bias due to culling. Mastitis was defined as a binary trait distinguishing between cows with at least one reported case during the defined period (1) and cows without cases (0) (**MAST**).

For SA, the observation of a cow with a case of mastitis was considered as uncensored, and time to first mastitis was defined as the number of days from 10 d before calving to the day of the first treatment of mastitis or culling due to mastitis (TFM). For a healthy cow, i.e. a cow without a case of mastitis, the observation was considered as right censored. For these cows the time period was defined as the number of days from 10 d before calving to, in named order, the day of next calving, the day of culling for other reasons than mastitis, the day of movement to a new herd or to lactation day 240. The latter figure was based on that the risk of culling for a cow increases approximately around that time after calving. The time period was restricted to a maximum of 700 d.

Statistical Analysis

As a preliminary analysis for SA, a Cox proportional hazard model (semi-parametric) was run and Kaplan-Meier curves created in order to check if data followed the Weibull distribution and thus, the adequacy of applying a Weibull proportional hazard model (fully parametric). The assumption was assessed graphically from plots of logs of the baseline survivor function, $\ln[-\ln S(t)]$ against logs of time, $\ln t$. If the Weibull assumption holds, the curve should be linear.

Two different statistical models were used in the main analysis: a linear model based on binary data and a Weibull proportional hazard model. For a better comparison of the 2 methods, the same fixed and random effects were used in both models. Each parity was analyzed separately for both models. For the mixed LM analysis the following sire model was used to analyze MAST:

$$y_{ijklm} = ym_i + age_j + hy_k + s_l + b_1 Het_m + b_2 Hol_m + e_{ijklm} \quad [1]$$

where y_{ijklm} is the observation of mastitis of cow m; ym_i is the fixed effect of i^{th} year by month at calving; age_j is the fixed effect of j^{th} age in months at calving (one month per class); hy_k is the random effect of k^{th} herd by year of calving; s_l is the random effect of l^{th} sire; b_1 is the fixed regression coefficient on proportion heterosis of animal m (Het_m); b_2 is the fixed regression coefficient on proportion North American Holstein of animal m (Hol_m); and e_{ijklm} is the random residual effect. Random effects were assumed normally distributed with zero means and variances $\mathbf{I}\sigma^2_{hy}$, $\mathbf{A}\sigma^2_s$ and $\mathbf{I}\sigma^2_e$, respectively, where **A** is the additive relationship matrix and **I** is the identity matrix.

The DMU package (Madsen and Jensen, 2000) was used to obtain REML estimates of the variance components. The heritability was calculated as:

$$h^2 = 4\sigma_s^2 \Big/ \left(\sigma_s^2 + \sigma_e^2\right) \quad [2]$$

The accuracy in selection was calculated according to:

$$r_{TI} = \sqrt{\frac{n}{n+k}} \quad [3]$$

where *n* is the number of daughters and $k = \frac{4 - h^2}{h^2}$

For the SA the following Weibull proportional hazard model was used to analyze TFM:

$$\lambda_{ijkl}(t) = \lambda_0(t) \exp\left\{ym_i + age_j + hy_k + s_l + b_1 Het_l + b_2 Hol_l\right\} \quad [4]$$

where $\lambda_{ijkl}(t)$ is the hazard of a cow getting mastitis at time t given that it has not occurred prior to t and $\lambda_0(t)$ the Weibull baseline hazard function $\left(\lambda\,\rho(\lambda t)^{\rho-1}\right)$ with scale parameter λ and shape parameter ρ. A value of ρ < 1 indicates that the hazard decreases with time, whereas ρ > 1 means that the hazard increases with time. The other effects, all time-independent, are as described for model [1]. The herd-year effect was assumed to follow a log-gamma distribution and was integrated out from the joint posterior density. The Weibull parameter ρ was estimated in the analyses.

Survival Kit V3.12 (Ducrocq and Sölkner, 2000) was used to estimate variance components for sire and herd-year. The heritability was calculated as:

$$h^2 = 4\sigma_s^2 \Big/ \left[\sigma_s^2 + (1/p)\right] \quad [5]$$

where *p* is the proportion of uncensored records.

This derivation for the heritability, which is measured on the original scale and not dependent on the Weibull parameters, was suggested by Yazdi et al. (2002) as the equivalent heritability. The accuracy was calculated as in [3].

RESULTS AND DISCUSSION

Descriptive Statistics

Results from the preliminary Cox analysis showed curves that were approximately linear after 3 to 5 days into lactation. Prior to that time point the curves were nonlinear due to the much higher frequency of mastitis around calving (results not shown). A way to handle this could have been to include stage of lactation in the model when running the Weibull analysis. However, in the present study the aim is to compare similar models for SA and LM, and therefore stage of lactation is not included.

The incidence of MAST increased with parity, with mean values of 10, 12 and 15 % for lactation 1 to 3. In the SA, the proportion of uncensored records, i.e. cows with mastitis, was about 15, 18, and 22 % for lactation 1 to 3, respectively.

Estimates of the Weibull Parameter

The estimated value of the Weibull parameter ρ varied between 0.57 and 0.70, which indicates that the risk of getting mastitis decreased with time within lactation. In our data, about 20-30% of all mastitis cases occurred before day 10 in lactation, with the highest values for first parity cows. If a time-dependent lactation stage effect was fitted, that effect would be expected to account for the early high risk, and the ρ would be expected to increase.

Heritabilities and Accuracies

Estimates of heritabilities and accuracies for MAST and TFM are provided in Table 2. The heritabilities of MAST (0.032 and 0.014 for first and later lactations, respectively) are in the range of reported estimates from other studies using LM. In a review by Heringstad et al. (2000) estimates of heritabilities of clinical mastitis from 13 studies based on Nordic data were between 0.001 and 0.06, with most values in the interval 0.02 to 0.03. Other estimates reported for first lactation range from 0.02 to 0.06 (Rupp and Boichard, 1999; Hansen et al., 2002; Carlén et al., 2004). Few studies have taken later parities into account and results are inconsistent. Pösö and Mäntysaari (1996) found higher heritabilities for lactation 2 and 3 in comparison with lactation 1, whereas Nielsen et al. (1997) did not find any differences in estimates between lactations. Heritability estimates on the linear scale are however influenced by frequency level and estimates from different studies are therefore not easily comparable (Emanuelson, 1988; Heringstad et al., 2000).

The heritability for TFM varied between 0.027 and 0.036 for lactation 1 to 3, with decreasing value with increasing parity. Since the proportions of uncensored records are considered in the calculations, these estimates can be compared with the heritability estimates for MAST. That the estimated heritabilities for TFM were higher than those estimated for MAST, is probably partly due to an increased observed variation among cows (TFM is a more continuously distributed trait than MAST). To our knowledge there are no previous studies where SA has been used to estimate the heritability of clinical mastitis.

In our study the heritability for both traits decreased with increasing lactation number. This was mainly an effect of increasing residual variances for MAST and decreasing sire variances for both MAST and TFM. The lower heritability for later lactations could also partly be explained by culling in first (and second) parity.

The accuracy of predicted breeding values for the first lactation was only slightly higher for TFM (0.76) than for MAST (0.74), whereas it was considerably higher for later lactations. Based on the increased accuracies, a gain in genetic progress would be expected by analyzing mastitis with SA instead of with LM, especially for later parities.

Although the difference in accuracy between TFM and MAST was small in first lactation, TFM can still be considered a more suitable trait. First of all, bias should be reduced since cows that are culled for other reasons than mastitis before they got a chance to express mastitis, no longer are considered as healthy cows, which they are if included when the trait MAST is analyzed. These other culling reasons may not be completely uncorrelated to mastitis and could therefore introduce a bias. For example, a cow culled due to another disease 10 d after calving is assigned 0 if the trait MAST is analyzed with LM and thus it is considered as good as a cow not getting mastitis within the whole study period (e.g. 150 d). If this cow had survived it might have contracted mastitis on day 11 after calving. When TFM is analyzed with SA, this cow will be censored with the censoring time 20 d.

Further, less information is lost with TFM. This is partly due to that cows with a case of mastitis get different TFM values, and so do cows without cases. Also, when MAST is analyzed in the traditional way and data is cut off (end of data collection), cows with less than for example 150 d of lactation and no reported mastitis case are either assigned 0, thus considered healthy, or their observations are treated as missing. With TFM these observations instead become censored and the information up to the censoring is utilized. Moreover, with the trait definitions used in this study, a cow getting mastitis 151 d after calving is assigned 0 and considered healthy when MAST is analyzed, whereas the same cow would be uncensored and treated as diseased with TFM. The fact that TFM, although only considering the first mastitis case, include cases after 150 d and is a more continuously distributed variable, might partly explain its higher heritability.

A feature to point out with TFM is that there will be variation in the accuracy of sire breeding values due to the proportion of uncensored daughters for each sire. Sires with a large proportion of daughters with mastitis will get more accurate breeding values. For first lactation in our data the percentage of daughters with mastitis (uncensored records) varied between 0 and 34, with mean 15.

Since TFM is a more continuous trait and distinguishes between cows within the healthy group, and within the affected group, one might wonder why this trait could not be analyzed with LM instead of with SA. However, the use of TFM is intimately linked to the possibility to handle censoring. Otherwise,

culled animals would have to be given a TFM of the culling time, even though they did not have mastitis. Likewise, cows not having mastitis but a next calving would get a TFM corresponding to that period. Both of these cows would be considered to be as bad as cows that actually had contracted mastitis at these times, a feature that is clearly undesirable. In SA, on the other hand, healthy and sick animals are clearly distinguished between.

Correlations Between Predicted Breeding Values

Correlations between breeding values predicted with LM and SA were 0.93, 0.89 and 0.88 for lactation 1 to 3, respectively. This implies that re-ranking among bulls occurred when the different methods were used. The somewhat lower correlations for later lactations correspond to the decreasing heritabilities and accuracies with increasing parity for both traits.

CONCLUSIONS

The main advantages of survival analysis over linear models when analyzing clinical mastitis are that more of the available information is utilized, thus increasing precision, and that culled cows are treated appropriately, thus reducing potential bias. The heritability estimates of time to first mastitis case analyzed with survival analysis were somewhat higher than the heritability estimates of clinical mastitis (0/1) analyzed with a linear model. The corresponding accuracies in selection were also higher for the trait time to first mastitis, especially for later lactations, and this may increase the genetic progress. More studies are needed before SA can be an alternative for genetic evaluations of clinical mastitis. It would for example be of interest to investigate inclusion of time-dependent effects in the model, e.g. stage of lactation. Nevertheless, the results indicate that survival analysis in the future can become a more suitable method than traditional linear model methodology to analyze clinical mastitis.

ACKNOWLEDGEMENTS

The authors are grateful to the Swedish Dairy Association for providing the data.

REFERENCES

Carlén, E., E. Strandberg and A. Roth. 2004. Genetic parameters for clinical mastitis, somatic cell score, and production in the first three lactations of Swedish Holstein cows. J. Dairy Sci. 87:3062-3070.

Ducrocq, V. 1987. An analysis of length of productive life in dairy cattle. Ph.D. Diss., Cornell Univ., Ithaca, NY.

Ducrocq, V. 1994. Statistical analysis of length of productive life for dairy cows of the Normande breed. J. Dairy Sci. 77: 855-866.

Ducrocq, V., and J. Sölkner. 2000. The survival kit V3.12 User´s manual. Version 3, release 12. http://www.boku.ac.at /nuwi/popgen.

Emanuelson, U. 1988. Recording of production diseases in cattle and possibilities for genetic improvements: A review. Livest. Prod. Sci. 20: 89-106.

Hansen, M., M. S. Lund, M. K. Sørensen, and L. G. Christensen. 2002. Genetic parameters of dairy character, protein yield, clinical mastitis, and other diseases in the Danish Holstein cattle. J. Dairy Sci. 85: 445-452.

Heringstad, B., G. Klemetsdal, and J. Ruane. 2000. Selection for mastitis resistance in dairy cattle: a review with focus on the situation in the Nordic countries. Livest. Prod. Sci. 64: 95-106.

Hirst, W. M., R. D. Murray, W. R. Ward, and N. P. French. 2002. A mixed-effects time-to-event analysis of the relationship between first-lactation lameness and subsequent lameness in dairy cows in the UK. Prev. Vet. Med. 54: 191-201.

Madsen, P., and J. Jensen. 2000. A users's guide to DMU. A package for analyzing multivariate mixed models. Version 6, release 4. Danish Institute of Agricultural Sciences, Denmark. 19 pp.

Nielsen, U. S., G. A. Pedersen, J. Pedersen, and J. Jensen. 1997. Genetic correlations among health traits in different lactations. Proceedings International workshop on genetic improvement of functional traits in cattle; health. Uppsala, Sweden, June, 1997. Interbull Bull. 15: 68-77.

Pösö, J., and E. A. Mäntysaari. 1996. Relationships between clinical mastitis, somatic cell score, and production for the first three lactations of Finnish Ayrshire. J. Dairy Sci. 79: 1284-1291.

Rupp, R., and D. Boichard. 1999. Genetic parameters for clinical mastitis, somatic cell score, production, udder type traits, and milking ease in first lactation Holsteins. J. Dairy Sci. 82: 2198-2204.

Saebo, S., T. Almoy, A. H. Aastveit, B. Heringstad, and G. Klemetsdal. 2002. Modelling time to first treatment of clinical mastitis as first passage times of stochastic processes. Proc. 7th World Congr. Genet. Appl. Livest. Prod., Montpellier, France CD-ROM communication N° 16-16.

Smith, S. P., and R. L. Quaas. 1984. Productive lifespan of bull progeny groups: failure time analysis. J. Dairy Sci. 67: 2999-3007.

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

Table 1. Structure of the data for the first 3 lactations of Swedish Holstein cows.

	Lactation 1	Lactation 2	Lactation 3
Number of lactations	221,104	122,280	59,233
Number of sires	838	784	673
Number of herd-year classes	31,511	22,023	13,570

Table 2. Heritabilities and accuracies of mastitis analyzed as a binary trait with linear models (LM) and as time to first mastitis with survival analysis (SA) for the first 3 lactations of Swedish Holstein cows. Daughter group size is assumed to be 150 for first parity, and with 75 % survival to next lactation.

	Heritability		Accuracy	
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