Use Of Transition Probabilities For Estimation Of Mastitis Resistance

J. Franzén[*], D. Thorburn[†], J. I. Urioste[*] and E. Strandberg[*]

Introduction

Mastitis is a common disease in dairy cattle with apparent economic consequences. Routine recording of clinical mastitis (CM) is not performed in most countries. Instead, somatic cell count (SCC) is commonly used in the genetic evaluation of mastitis, due to the genetic correlation between the two (Mark *et al.* 2002; Miglior *et al.* 2005). The degree of which the dynamic nature of the SCC lactation pattern is used in studies varies from modeling the lactation average to more comprehensive methods that better capture changes in SCC (Detilleux and Leroy 2000; de Haas *et al*. 2004 and 2008; Green *et al*. 2004; Madsen *et al*. 2008).

We present an alternative approach in which genetic evaluation of mastitis is performed by studying SCC changes during the lactation. These changes are captured by modeling transitions probabilities between assumed states of mastitis and non-mastitis. A widely dispersed SCC pattern will generate high transitions probabilities and should therefore be a good reflection of the SCC fluctuations over the lactation. The method will simultaneously model the transition probability of getting a case of mastitis and the probability of recovering from an infection. The former has been the focus of many studies but the infection recovery process has been scarcely investigated. In this approach we model both parts with the aim of capturing as much genetic information as possible from the SCC lactation pattern.

Material and methods

Data. Weekly SCC values were simulated for four populations with different combinations of population sizes and mastitis frequencies. Two population sizes were used; 24,000 and 60,000 first-parity cows. The cows were daughters of 400 unrelated sires distributed over 1,200 herds giving an average daughter group size of 60 and 150, respectively. For each population size, two different average mastitis frequencies were simulated (0.22 and 1.36 cases/lactation). Sire and cow breeding values for mastitis liability were simulated according to Carlén *et al.* (2006). From the underlying normally distributed liability, mastitis data were created as a weekly binary trait. SCC values for uninfected test days were then simulated as random deviations from a baseline curve and mastitis test days as random deviations from a function with instant SCC increase followed by a successive decline down to the baseline level (Hagnestam-Nielsen and Østergaard 2009).

Model and statistical analysis. We assume that a cow *i* during a lactation moves between two possible states: sick and healthy. A prespecified SCC level marks the threshold between

[*] Dept. of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala, Sweden
[†] Dept. of Statistics, Stockholm University, Stockholm Sweden

sick and healthy states. The threshold varies along the lactation according to a multiple of an average lactation curve for primiparous cows. The lactation curve was modeled by a spline function, parameterized according to findings of de Haas *et al*. (2002).

The probability of going from a healthy to a sick state is denoted $P_i^{(HS)}$ and the probability of a state change in the other direction, $P_i^{(SH)}$. The transition probabilities for a cow *i* may be summarized in a transition matrix T_i,

$$T_i = \begin{bmatrix} P_i^{(HS)} & 1-P_i^{(HS)} \\ 1-P_i^{(SH)} & P_i^{(SH)} \end{bmatrix}$$

which gives the probabilities of going between the two possible states or remain in a current state. A transition between two particular states is repeatable, meaning a cow may be in a particular state more than once during a given lactation. An episode is the time spent in each state; after a transition, a new episode begins, leading to multiple episodes within the lactation. When modeling survival data with repeated events, such as repeated mastitis cases during a lactation, multilevel models are powerful tools (Browne *et al*. 2009).

Data for mastitis and SCC are most often interval-censored, meaning the actual time for a transition between states is not known. When data is collected retrospectively, a state change is only known to have occurred sometime between two data collection times, referred to as time interval *t*. Due to the structure of data, a multilevel discrete time survival model is used as opposed to a continuous time model such as the Cox proportional hazards model. By expanding the data set, the model can be treated as a multilevel binary response model. Repeated transitions can be viewed as having a three-level hierarchical structure with episodes nested within cows. Herds and sires are cross classified on the highest level.

The probability P_{ijk} is the discrete correspondence to the continuous time hazard function and is defined as the probability that a transition occurs sometime between any two measurement occasions for cow *i*, daughter if sire *j*, and member of herd *k* . The model for the transition probability of going from a healthy to a sick state, $P_{ijk}^{(HS)}$ is expressed as:

$$y_{ijk(t)} \sim Ber\left(P_{ijk}^{(HS)}\right) \quad \text{and} \quad \text{logit}\left(P_{ijk}^{(HS)}\right) = \beta^{(HS)} + s_j^{(HS)} + h_k^{(HS)} + e_{ijk}^{(HS)}$$

where $y_{ijk(t)}$ takes value 1 if a transition occurred in time interval *t* and 0 otherwise. $s_j^{(HS)}$ and $h_k^{(HS)}$ are the random sire and herd effects and $e_{ijk}^{(HS)}$ is the random residual effect. The transition probability of going the other direction, $P_{ijk}^{(SH)}$, is modeled in a corresponding way.

Estimations were made in the multilevel software program MLwiN using Bayesian inference and MCMC simulations. Vague priors were used; the variances of the random effects have an inverse Gamma prior, $p\left(\sigma^2\right) \sim \Gamma^{-1}\left(\varepsilon, \varepsilon\right)$, $\varepsilon = 0.001$, and the fixed intercept has a uniform prior, $p\left(\beta\right) \propto 1$.

Results and discussion

Correlations between true and estimated breeding values ranged between 0.5 and 0.8 for the part of the model handling transitions from non-mastitis to mastitis states (Table 1). Correlations increased with increasing daughter group size, mastitis frequency and threshold level. The method is still to be evaluated for the transitions in the reverse direction. The major strength of the method is its ability to take into account repeated mastitis cases within lactations. This is shown through higher correlations for higher mastitis frequencies; a higher threshold level also gives better correlations. The optimal boundary between the states is a question that should be investigated further. A too low threshold will falsely label high random fluctuations around "normal" SCC levels as cases of mastitis. There is a trade-off between classifying normal high SCC values as infected if the level is too low and missing possible infections if the level is too high.

Table 1: Correlations between true and estimated breeding values for situations with different number of daughters per sire, mastitis frequencies and two levels of the SCC threshold between healthy and sick (5 or 10 times the base curve)

	Number of daughters			
	60		**150**	
	Mastitis frequency (cases/lact.)			
Threshold	**0.22**	**1.36**	**0.22**	**1.36**
5×BC	0.52	0.61	0.73	0.76
10×BC	0.56	0.70	0.73	0.82

The simulated data in this study has previously been used for evaluation of linear models (LM), threshold models (TM) and survival analysis (SA) (Carlén *et al.,* 2006). Confirmed cases of mastitis were modeled either as a binary variable or as the time to the first case. Estimated correlations in that study were 0.53-0.60 and 0.70-0.76 for 60 and 150 daughters, respectively. Thus, the results for the transition method compare well with those of Carlén *et al.* (2006). Considering that the transition method is using only SCC values as an indicator, the results are even more promising. In addition, there is probably more valuable information to gather from the results of analyzing transitions in the opposite direction.

When the number of cases per lactation increases, the new method is generating higher correlations than the result from LM, TM, and SA study; this is not surprising because the latter methods only take the first case into account. The lactation frequency of CM cases is usually rather low. If cases of subclinical mastitis (SCM) were to be included in the mastitis definition, the number of cases would increase. By lowering the threshold between states, these

cases could be included in the mastitis classification. Future studies will reveal if the method is also favorable for higher mastitis frequencies when added cases refer to SCM. Another possibility that the method gives is to add a new state representing subclinical cases. We would then model transition probabilities between three possible states.

Conclusion

The transition probability approach for modeling SCC pattern as an indication for CM gives promising results in the primary test run. Correlations between true and estimated breeding values are at least as high compared to three other methods which model confirmed cases of mastitis. This calls for further and more comprehensive investigations of the method.

Acknowledgements

This work was carried out as part of the RobustMilk project that is financially supported by the European Commission under the Seventh Research Framework Programme, Grant Agreement KBBE-211708. The content of this paper is the sole responsibility of the authors, and it does not necessarily represent the views of the Commission or its services.

References

Browne, W. J., Steele, F., Golalizadeh, M., *et al*. (2009), *J. R. Stat. Soc. Ser. A*, 172:579-598.

Carlén, E., Emanuelson, U., and Strandberg, E. (2006). *J. Dairy Sci*., 89:4049-4057.

De Haas Y., Barkema, H.W., and Veerkamp, R. (2002). *J. Dairy Sci*., 85:1314-1323.

De Haas, Y., Veerkamp, R. F., Barkema, H. W. *et al.* (2004). *J. Dairy Sci.* 87:95-105.

De Haas Y., Ouweltjes W., ten Napel J. *et al*. (2008). *J. Dairy Sci.* 91:2501-2511.

Detilleux, J., and Leroy, P. L. (2000). *J. Dairy Sci.* 83:2341-2349

Green, M. J. , Green, L. E. , Schukken, Y. H. *et al.* (2004). *J. Dairy Sci.* 87:1256-1264.

Hagnestam-Nielsen, C., and Østergaard, S. (2009). *Animal*, 3:315-328.

Madsen, P., Shariati, M. M., and Ødegård, P. (2008). *J. Dairy Sci.* 91:4355-4364.

Mark, T., Fikse, W. F., Emanuelson, U., *et al.* (2002). *J. Dairy Sci.* 85:2384–2392.

Miglior, F., Muir, B.L., and Van Doormaal, B. J. (2005). *J. Dairy Sci.*, 88:1255–1263