Longitudinal Analysis of Somatic Cell Count for Joint Genetic Evaluation of Mastitis and Recovery Liability

B.G. Welderufael [*†], **D.J. De Koning** [*], **L. Janss**[†], **J. Franzén** [‡] **and W.F. Fikse**[*]
[*]Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala, Sweden.
[†] Center for Quantitative Genetics and Genomics, Department of Molecular Biology and Genetics, Aarhus University, Tjele, Denmark. [‡]Department of Statistics Stockholm University, Stockholm, Sweden.

ABSTRACT: Better models of genetic evaluation for mastitis can be developed through longitudinal analysis of somatic cell count (SCC) which usually is used as a proxy for mastitis. Mastitis and recovery data with weekly observations of SCC were simulated for daughter groups of 60 and 240 per sire. Data were created to define cases: 1 if SCC was above a pre-specified boundary, else 0. A transition from below to above the boundary indicates probability to contract mastitis, and the other way indicates recovery. The MCMCglmm package was used to estimate breeding values. In the 60 daughters group, accuracies ranged from 0.53 to 0.54 for mastitis and 0.22 to 0.23 for recovery. Whereas, in the 240 daughters group accuracies ranged from 0.83 to 0.85 for mastitis and 0.57 to 0.65 for recovery. Reasonable accuracies can be achieved from SCC based estimates.

Keywords: Mastitis liability; Recovery liability; Somatic cell count

Introduction

Mastitis, an inflammation of the mammary gland, is usually caused by a bacterial infection. Occasionally, it might arise from chemical, mechanical, or thermal injuries. It is a common and costly disease in modern dairy farms. In addition to its economic importance (due to discarded milk, reduced milk production, culling cows and treatments cost), dairy cows welfare and consumers' demand to antibiotics-free milk and milk products are among the main mastitis-related threats to the modern dairy farms. In addition to this the well documented (Carlén et al. (2006); Hinrichs et al. (2005)) unfavorable genetic correlation between milk production and mastitis calls for the inclusion of mastitis in a dairy cattle breeding program. Better modelling and genetic evaluation of mastitis are a requirement for breeding mastitis free and/or resistant dairy cows.

Usually genetic evaluation of mastitis liability is performed either with cross-sectional or longitudinal approaches (Franzén et al. (2012)). In cross-sectional models, lactations are considered as a static process and data is recorded when a mastitis event is observed while mastitis itself is a developmental process. In longitudinal models, developments that occur throughout a lactation may have to be considered by the model (Franzén et al. (2012)). Therefore, longitudinal modelling captures as much genetic information as possible by analyzing the changes or developments that each cow experiences during lactation. According to Franzén et al. (2012), most of the previous genetic evaluations of mastitis focused only on one direction of the disease: mastitis susceptibility, thereby ignoring the recovery process.

In the genetic evaluations of mastitis, the difficulty or lack of routine records (Carlén et al. (2006)) of the most common type of mastitis (subclinical mastitis) cases has led to the use of indirect but related measures, such as: somatic cell counts (SCC), udder conformation and others. Among these related traits SCC is widely used as a proxy for mastitis (Uhler 2009) due to its higher correlation with clinical mastitis. In addition, SCC is unfavorably correlated with milk production and consequently cow welfare.

Most of the recent genetic analyses using SCC data (Franzén et al. (2012)) focused on single-trait models, ignoring the possible genetic correlation between the liability of becoming sick and the recovery process. In the present study we perform a multi-trait sire model longitudinal genetic analysis to evaluate the level of accuracy achieved from SCC based estimates compared to the simulated mastitis liability and recovery. Accuracies were also compared on different daughter groups per sire and level of mastitis incidence (cases per lactation).

Materials and Methods

Data. Simulation software by Carlén et al. (2006) and Franzén et al. (2012) was used to generate records on milk production, interval between calving and first ovulation, conception liability, mastitis liability and recovery.

A mastitis history for the whole lactation was generated on the basis of liability for contracting and recovering from mastitis. Weekly SCC values were generated on the basis of this mastitis and recovery history. SCC on non-mastitic test-days followed a baseline lactation curve, to which random noise was added. The level of SCC was elevated on mastitic test-days.

Two level of mastitis incidence or cases per lactation (28% and 95%) and three different genetic correlations ($r_g = 0.0$, $r_g = 0.2$ and $r_g = -0.2$) between mastitis liability and recovery were considered. Five replicates were used for each case. The cows were daughters of 100 and 400 unrelated sires distributed over 1200 herds resulted into two daughter group sizes of 240 and 60 per sire. More details on the parameters used to simulate the data sets can be found in Franzén et al. (2012).

Binary data were created based on the generated weekly SCC observations to define whether a cow was in a disease (*D*) or healthy (*H*) state. The SCC based boundary ($B(\tau)$) was allowed to vary along the lactation curve ($L(\tau)$)

according to this multiplier: $B(\tau) = m \times L(\tau)$; where τ is time in lactation, starting at calving. If a cow's weekly SCC exceeded this boundary then the cow was considered to have mastitis. The multiplier factor (m = 10) used to create the binary data lowers chances of misclassifications (Franzén et al. (2012)). If a cow's weekly SCC was larger than 200000 cells/mL then the cow was considered mastitic.

Transition probability models. For a cow in lactation, there is a possibility of contracting mastitis and recovery. A cow may contract mastitis and recover and vice versa according to the SCC level so that transition from *H* to *D* (tranHD) and from *D* to *H* (tranDH) is possible within a given lactation. This phenomenon was modeled with transition probability model, T_i, that shows the transition probabilities for individual cow *i* going from *H* to *D* state or from *D* to *H* or remain in a current state.

$$T_i = \begin{bmatrix} \pi^{(\text{tranHD})} & 1-\pi^{(\text{tranHD})} \\ 1-\pi^{(\text{tranDH})} & \pi^{(\text{tranDH})} \end{bmatrix}$$

Where:

$\pi^{(\text{tranHD})}$ = Probability of moving from a H to a D state

$\pi^{(\text{tranDH})}$ = probability of moving from a D to a H state

$1-\pi^{(\text{tranHD})}$ = probability of remaining in the H state

$1-\pi^{(\text{tranDH})}$) = probability of remaining in the D state

Thus, the sequence of *H*'s and *D*'s indicating whether or not a cow had mastitis on subsequent test days was converted into a new sequence of state changes: 0 if a cow remained in the same state and 1 if the cow changed state.

Statistical modelling. A multi-trait sire model was fitted to the binary data of transitions. The analyses were simplified by considering the transition data as normally distributed traits. Thus, the transition probability π_{ijk} which is defined as the probability that a transition occurs for cow *i*, daughter of sire *j* for an observation *k* was modelled as:

$$Y_{ijk} = \beta + C_i + S_j + e_{ijk}$$

Where:

Y_{ijk} = 1 if a transition occurs, otherwise 0.

β = liability of mastitis or recovery for an average cow

C_i = random sire effect

S_j = random cow effect

e_{ijk} = random residual effect for cow *i*

Analysis, Sampling and Bayesian inference. The analysis was performed in R (R Development Core Team 2013) with the MCMCglmm package (Hadfield 2010). The package implements Markov chain Monte Carlo (MCMC) routines for fitting multi-response generalized linear mixed models.

The MCMCglmm was run with its default values of 13000 iterations, 3000 burn-in and a thinning interval of 10. Point estimates of parameters were derived from the samples of the posterior distribution. Before considering the posterior estimates in further analysis, convergence diagnosis methods (trace plot, density plot and cross-corr) from the coda package (Martyn et.al. (2006)) were performed to evaluate the model fitness and sampling behavior of the MCMC procedure.

Breeding values were estimated for the two transition directions (tranHD and tranDH). Correlations between true breeding values (TBV), generated from the simulation process, and estimated breeding values (EBV) from the MCMCglmm analysis were calculated as the accuracy in both directions. The multi-trait model analysis provides an option to estimate the genetic correlations ($\hat{r}_g$) between the two transition directions.

Results and Discussion

Sampling and independency. Results from the MCMC chains were analyzed using the coda package in R. The trace plot and posterior distributions of the variance components (sire variance and covariance) showed that the MCMC algorithm was well converged (Figure 1). There are some autocorrelations in the sampling process of the MCMC as trends are apparent in the top left-hand and bottom right-hand of the plot (Figure 1).

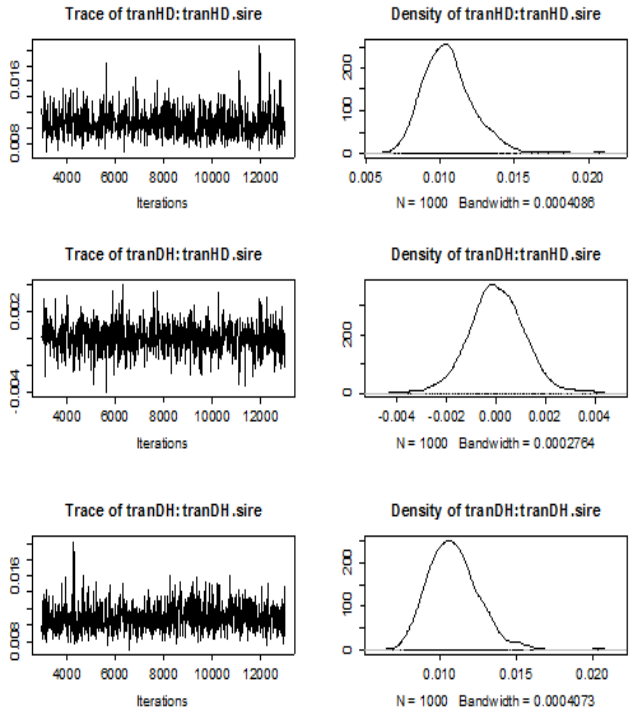

Figure 1. Markov chain Monte Carlo trace and density plots of sire variance and covariance between the two directions (tranHD, tranDH).

Estimation of accuracies. The estimation accuracies (correlations between TBV and EBV) in the smaller daughters group per sire ranged from 0.53 to 0.54 and from 0.22 to 0.23 for mastitis contract and recovery

respectively (Table 1). For larger daughters group per sire the accuracies ranged from 0.83 to 0.85 for mastitis liabilities and from 0.57 to 0.65 for the recovery process. These estimates are reasonably precise as the standard errors in all cases ranged from 0.00 to 0.03. In earlier studies by Franzén et al. (2012) higher accuracies of EBV were reported for larger daughter group size (150) coupled with higher mastitis frequencies. Accuracies were more influenced by daughters group per sire than by mastitis cases per lactation (results not presented).

Table 1. Average correlation between true breeding value (TBV) and estimated breeding value (EBV).

Daughters per sire	60			240		
r_g	HD[1]	DH[2]	$\hat{r}_g$[3]	HD	DH	$\hat{r}_g$
$r_g = 0.0$	**0.53**	**0.23**	0.00	**0.85**	**0.62**	0.00
	0.01	0.00	0.00	0.01	0.03	0.00
$r_g = 0.2$	**0.53**	**0.22**	0.00	**0.83**	**0.57**	0.00
	0.02	0.02	0.00	0.02	0.02	0.00
$r_g = -0.2$	**0.54**	**0.23**	0.00	**0.85**	**0.65**	0.00
	0.02	0.01	0.00	0.01	0.01	0.00

Figures in bold represent average correlation between TBV and EBV of 5 different replicates. Under each average correlation is the standard error of estimate for these replicates.
[1]HD = mastitis liability; [2]DH = recovery process.
[3]$\hat{r}_g$ = estimated genetic correlations between HD and DH.

Our estimations of accuracies were not dependent on the simulated values of possible genetic correlations between mastitis and recovery liabilities ($r_g = 0.0$, $r_g = 0.2$ and $r_g = -0.2$). Furthermore, the estimated genetic correlations did not show any trend with these simulated correlations. The average estimated genetic correlation (Table 1) was approximately zero ($\hat{r}_g \approx 0$).

For the disease to health direction the information in the data was much lower. Because there were large number of cows without mastitis (small number of tranDH) that resulted into poor estimates (0.25) of accuracies (Franzén et al. 2012). In the present study we have achieved higher accuracies (0.65) by increasing the daughter groups per sire while keeping the mastitis frequencies the same. The larger daughters group represented in this study is not uncommon in the modern dairy farms.

Conclusion

The model generates higher accuracies for the health to disease transitions for both daughter group sizes. The more daughters per sire, the more accurate the estimates were. For the disease to health direction achieving higher accuracies with the smaller daughter group size was not possible. However the model gives an option to include the two direction of the disease in the analysis. It also demonstrated the importance SCC in the genetic evaluation of mastitis and recovery with reasonably higher accuracies.

Acknowledgements

BGW benefited from a joint grant from the European Commission and SLU, within the framework of the Erasmus-Mundus joint doctorate "EGS-ABG".

Literature Cited

Browne, W. J., and Draper, D. (2006). *Bayesian Anal.*, 1: 473-513.

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

Franzén, J., Thorburn, D., Urioste, J. I. et. al. (2012). *Genet Select Evol.*, 44:10.

Hadfield, J. D. (2010). *J. Stat. Softw.*, 33: 1-22.

Hinrichs, D., Stamer, E., Junge, W. et. al. (2005). *J. Dairy Sci.*, 88: 2260-2268

Martyn, P., Nicky, B., Kate, C. et.al. (2006). *R News*, 6:7-11.

R Development Core Team. (2013). http://www.R–project.org/.

Uhler, C. (2009). *J. ABES*, 14:79-98.