Analysis of Test Day Somatic Cell Score Using a Liability-Normal Mixture Model

P. Madsen[1] and J. Ødegård[2]
[1]*Danish Institute of Agricultural Sciences, Dept. of Genetics and Biotechnology, Foulum, Denmark;*
[2]*Norwegian University of Life Sciences, Dept. of Animal and Aquacultural Sciences, Ås, Norway*

Introduction

Breeding programs utilizing somatic cell count (**SCC**) or somatic cell score (**SCS**) records in selection for improved udder health have so far been based on either cross-sectional models (e.g., lactation mean SCS) or test-day models (e.g., repeatability models, random regression models). All these models focus on selecting animals with the lowest average SCS. However, selection for lower average SCS might not only favor animals with low incidence of intramammary infections (**IMI**), but also animals having low levels of SCC when healthy ("baseline" SCC). Further, it has been argued that a high "baseline" SCC may improve the cows' resistance to IMI (Detilleux and Leroy, 2000; Gianola *et al.,* 2004).

Observed test-day SCS in the milk of a cow can be viewed as a variable drawn from two (or more) distributions, depending on the udder health status of the cow on the day of sampling. In its simplest case test-day udder health status may be categorized as either with or without IMI (i.e., "mastitic" or "healthy"). Hence, test-day SCS records represent a mixture of at least two distributions depending on the IMI status on the test-day. These distributions may differ with respect to expected values and possibly also with respect to dispersion parameters. In field data, the IMI status on a test-day is typically unknown

Two-component normal mixture models for SCS have been developed, using either Maximum Likelihood (Detilleux and Leroy, 2000; Gianola *et al.,* 2004) or Bayesian techniques (Ødegård *et al.,* 2003). In these models, SCS was assumed to be normally distributed; with either homogeneous (Gianola *et al.,* 2004) or heterogeneous residual variances (Ødegård *et al.,* 2003), and with location parameters including both systematic and random effects. The expected increase in test-day SCS, as a response to IMI, was accounted for by including the effect of putative IMI status among the location parameters for SCS. However, these models assumed identical *a priori* probabilities of IMI for all observations, and they did not provide a sufficient tool for selecting animals for lower probability of mastitis, given the observed SCS, as all genetic effects are calculated for SCS level, adjusted for mastitis effects. A more realistic liability normal mixture (**LNM**) model was developed by Ødegård *et al.* (2005), where the udder health status (which is an unobserved binary variable) is assumed fully determined by an unobserved underlying liability. Location parameters for the liability can include both systematic and random effects, allowing the probability of IMI to differ between animals as well as between observations within animal. The LNM model predicts genetic effects for both SCS and for the unobserved liability to IMI, where the latter can be used in selection for improved udder health. As previously stated, genetic level of "baseline" SCS, may have some relevance for risk of subsequent infection, and may therefore contain valuable information for genetic improvement of udder health. Such potentially valuable information was utilized by including a covariance structure between the random effects for SCS and liability to IMI in the model.

The objective of this study was to analyze real test-day SCS data with a LNM-model and examine its suitability for use in practical selection for improved resistance to mastitis.

Material and Methods

Method. Setting and notation are as in Ødegård *et al.* (2005). Briefly, the data consisted of *n* measurements for SCS. A two-component normal mixture model poses that

the vector SCS, consisting of all SCS observation, given some location and dispersion parameters ($\boldsymbol{\alpha}$), and probabilities (**P**) have the mixture density (assuming conditional independence):

$$p(\mathbf{SCS}|\mathbf{P},\boldsymbol{\alpha}) = \prod_{i=1}^{n}\left[N^{*}\left(f_i^{*}(\boldsymbol{\alpha}),g_i^{*}(\boldsymbol{\alpha})\right)(1-P_i)+N\left(f_i(\boldsymbol{\alpha}),g_i(\boldsymbol{\alpha})\right)P_i\right] \quad [1]$$

where P_i is the *a priori* probability of IMI+ for observation *i*. Typically, $f_i^{*}(\boldsymbol{\alpha})$ and $f_i(\boldsymbol{\alpha})$ are linear combinations of fixed and random effects, and $g_i^{*}(\boldsymbol{\alpha})=\sigma^2_{e0_{SCS}}$ and $g_i(\boldsymbol{\alpha})=\sigma^2_{e1_{SCS}}$ are variance parameters. Estimation is facilitated by augmenting the density above with auxiliary binary indicator variables Z_i (IMI- $\rightarrow$ 0, IMI+ $\rightarrow$ 1). An underlying continuous random variable is assumed, called liability (λ), which determines the mastitis status (Z_i) associated with each observation, depending on the value of liability relative to a fixed threshold (0). Thus, P_i can be written as:

$$P_i = Pr(\lambda_i > 0|\boldsymbol{\alpha}), \quad \text{so} \quad \text{that: } Z_i = \begin{cases} 0 \textbf{ if } \lambda_i \le 0 \\ 1 \textbf{ if } \lambda_i > 0 \end{cases} \quad [2]$$

It is further assumed that:

$$\mathrm{p}(\boldsymbol{\lambda},\mathbf{SCS}|\mathbf{Z}=\mathbf{z},\boldsymbol{\alpha}) = \mathrm{p}(\boldsymbol{\lambda}|\mathbf{Z}=\mathbf{z},\boldsymbol{\alpha})\mathrm{p}(\mathbf{SCS}|\mathbf{Z}=\mathbf{z},\boldsymbol{\alpha}) \quad [3]$$

implying that the residual correlation between SCS (adjusted for IMI effects) and liability is zero.

Let

$$\boldsymbol{\alpha} = \left(\boldsymbol{\beta},\mathbf{h},\mathbf{a},\mathbf{p},\mathbf{H_0},\mathbf{G_0},\mathbf{P_0},\sigma^2_{e\mathbf{0}_{SCS}},\sigma^2_{e\mathbf{1}_{SCS}}\right),$$

where $\boldsymbol{\beta}=\left[\boldsymbol{\beta}_{\mathbf{0}_{SCS}}' \quad \boldsymbol{\beta}_{\mathbf{1}_{SCS}}' \quad \boldsymbol{\beta}_{\lambda}'\right]'$, $\mathbf{h}=\left[\mathbf{h}_{SCS}' \quad \mathbf{h}_{\lambda}'\right]'$, $\mathbf{a}=\left[\mathbf{a}_{\mathbf{0}_{SCS}}' \quad \mathbf{a}_{\mathbf{1}_{SCS}}' \quad \mathbf{a}_{\lambda}'\right]'$, and $\mathbf{p}=\left[\mathbf{p}_{\mathbf{0}_{SCS}}' \quad \mathbf{p}_{\mathbf{1}_{SCS}}' \quad \mathbf{p}_{\lambda}'\right]'$ are vectors of "fixed" ($\boldsymbol{\beta}$), random herd-test-day (**h**), random additive genetic (**a**), and random permanent environmental (**p**) effects on SCS and liability to mastitis, where the sub-vectors $\boldsymbol{\beta}_{\mathbf{0}_{SCS}}$, $\mathbf{a}_{\mathbf{0}_{SCS}}$ and $\mathbf{p}_{\mathbf{0}_{SCS}}$ includes effects affecting SCS irrespective of IMI status, while $\boldsymbol{\beta}_{\mathbf{1}_{SCS}}$, $\mathbf{a}_{\mathbf{1}_{SCS}}$ and $\mathbf{p}_{\mathbf{1}_{SCS}}$ includes effects peculiar to SCS in cows with IMI+. Further, $\mathbf{H_0}$ is the 2 × 2 (co)variance matrix of herd-test-day effects, $\mathbf{G_0}$ is the 3 × 3 (co)variance matrix of additive genetic effects, and $\mathbf{P_0}$ is the 3 × 3 (co)variance matrix of permanent environmental effects; $\sigma^2_{e\mathbf{0}_{SCS}}$ and $\sigma^2_{e\mathbf{1}_{SCS}}$ are the residual variances of SCS in the IMI- and IMI+ classes, respectively. Given IMI status (**Z**) and $\boldsymbol{\alpha}$, SCS can be modeled as a Gaussian trait, allowing for heterogeneous (co)variance components for the two disease categories. The density of the conditional distribution of **SCS**, given **Z** = **z** and $\boldsymbol{\alpha}$ is:

$$p(\mathbf{SCS}|\boldsymbol{\alpha},\mathbf{Z}=\mathbf{z}) = \prod_{i=1}^{n} N\left((1-Z_i)f^{*}(\boldsymbol{\alpha})+Z_i f(\boldsymbol{\alpha}),(1-Z_i)\sigma^2_{e0_{SCS}}+Z_i\sigma^2_{e1_{SCS}}\right) \quad [4]$$

where

$$f^{*}(\boldsymbol{\alpha}) = \left(\mathbf{x}'_{\mathbf{i0}_{SCS}}\boldsymbol{\beta}_{\mathbf{0}_{SCS}}+\mathbf{w}'_{\mathbf{ih}}\mathbf{h}_{SCS}+\mathbf{w}'_{\mathbf{ia}}\mathbf{a}_{\mathbf{0}_{SCS}}+\mathbf{w}'_{\mathbf{ip}}\mathbf{p}_{\mathbf{0}_{SCS}}\right)$$

$$, f(\boldsymbol{\alpha}) = \left(f^{*}(\boldsymbol{\alpha})+\mathbf{x}'_{\mathbf{i1}_{SCS}}\boldsymbol{\beta}_{\mathbf{1}_{SCS}}+\mathbf{w}'_{\mathbf{ia}}\mathbf{a}_{\mathbf{1}_{SCS}}+\mathbf{w}'_{\mathbf{ip}}\mathbf{p}_{\mathbf{1}_{SCS}}\right),$$

and **x**′ and **w**′ with appropriate subscripts are incidence row vectors. Further, it is assumed that the density of the $\boldsymbol{\lambda}$ vector, given $\boldsymbol{\alpha}$ is:

$$p(\boldsymbol{\lambda}|\boldsymbol{\alpha}) = \prod_{i=1}^{n} N\left(\mathbf{x}'_{\mathbf{i}_{\lambda}}\boldsymbol{\beta}_{\lambda}+\mathbf{w}'_{\mathbf{ih}}\mathbf{h}_{\lambda}+\mathbf{w}'_{\mathbf{ia}}\mathbf{a}_{\lambda}+\mathbf{w}'_{\mathbf{ip}}\mathbf{p}_{\lambda},1\right) \quad [5]$$

With this structure [2] is equivalent to

$$P_i = \Phi\left(\mathbf{x}_{\mathbf{i}_{\lambda}}'\boldsymbol{\beta}_{\lambda}+\mathbf{w}'_{\mathbf{ih}}\mathbf{h}_{\lambda}+\mathbf{w}_{\mathbf{ia}}'\mathbf{a}_{\lambda}+\mathbf{w}_{\mathbf{ip}}'\mathbf{p}_{\lambda}\right) = \Phi\left(\tilde{\lambda}_i\right) \quad [6]$$

Conditional density of SCS. Given [3] conditional density of SCS is:

$$p(\mathbf{SCS}|\boldsymbol{\alpha},\boldsymbol{\lambda},\mathbf{Z}=\mathbf{z}) = p(\mathbf{SCS}|\boldsymbol{\alpha},\mathbf{Z}=\mathbf{z}) = [4].$$

Prior density of all unknown parameters. The joint prior density of all unknown parameters, including the liabilities ($\boldsymbol{\lambda}$) as unknowns, is:

$$p(\boldsymbol{\alpha},\boldsymbol{\lambda},\mathbf{Z}=\mathbf{z}) = \Pr(\mathbf{Z}=\mathbf{z}|\boldsymbol{\alpha},\boldsymbol{\lambda})p(\boldsymbol{\lambda}|\boldsymbol{\alpha})p(\boldsymbol{\alpha})$$
$$= \Pr(\mathbf{Z}=\mathbf{z}|\boldsymbol{\lambda})\times[5]\times p(\boldsymbol{\alpha}) \qquad [7]$$

Given $\boldsymbol{\lambda}$, $\mathbf{Z}$ is completely specified, and $\Pr(\mathbf{Z}=\mathbf{z}|\boldsymbol{\lambda})$ is, therefore, a degenerate distribution, and the density $p(\boldsymbol{\alpha})$ is

$$p(\boldsymbol{\alpha}) = p(\boldsymbol{\beta})p(\mathbf{h}|\mathbf{H_0})p(\mathbf{a}|\mathbf{G_0})p(\mathbf{p}|\mathbf{P_0})p(\mathbf{H_o})$$
$$p(\mathbf{G_o})p(\mathbf{P_0})p(\sigma^2_{e0_{SCS}})p(\sigma^2_{e1_{SCS}}) \qquad [8]$$

where $p(\boldsymbol{\beta})$, $p(\mathbf{H_o})$, $p(\mathbf{G_o})$, $p(\mathbf{P_0})$, $p(\sigma^2_{e0_{SCS}})$, $p(\sigma^2_{e1_{SCS}})$ were assigned bounded uniform priors. To achieve reasonably vague priors, the absolute values of the bounds were large. Further, to avoid "label-switching" problems (Mclachlan and Peel, 2000), constraints were imposed on parameters of the SCS distributions of putative IMI- and IMI+ animals. Herd-test-day effects (**h**), additive breeding values (**a**) and permanent environmental effects (**p**) were assumed to be normally distributed, using standard assumptions.

Joint posterior density. The augmented joint posterior density of all unknowns is:

$$p(\boldsymbol{\alpha},\boldsymbol{\lambda},\mathbf{Z}=\mathbf{z}|\mathbf{SCS}) \propto$$
$$p(\mathbf{SCS}|\boldsymbol{\alpha},\boldsymbol{\lambda},\mathbf{Z}=\mathbf{z})p(\boldsymbol{\alpha},\boldsymbol{\lambda},\mathbf{Z}=\mathbf{z}) = [4]\times[7] \qquad [9]$$

Fully conditional posterior distributions. Given $\mathbf{Z}$ and $\boldsymbol{\lambda}$, all fully conditional posterior distributions for location and dispersion parameters have standard forms as in a linear Gaussian model. The distributions of elements from the vector Z, conditional on SCS and the parameter vector $\boldsymbol{\alpha}$ (unconditional on the liability) have Bernoulli distributions with parameters τ_i, equaling:

$$\tau_i = \frac{p(SCS_i|\boldsymbol{\alpha},Z_i=1)\Phi(\tilde{\lambda}_i)}{p(SCS_i|\boldsymbol{\alpha},Z_i=0)(1-\Phi(\tilde{\lambda}_i))+p(SCS_i|\boldsymbol{\alpha},Z_i=1)\Phi(\tilde{\lambda}_i)} \qquad [10]$$

Given Z_i, λ_i have a distribution with density;

$$p(\lambda_i|\boldsymbol{\alpha},Z_i=z_i) = (1-Z_i)p(\lambda_i|\boldsymbol{\alpha},Z_i=0)+Z_i p(\lambda_i|\boldsymbol{\alpha},Z_i=1). \qquad [11]$$

The fully conditional distribution of λ_i, given Z_i, is a truncated standard normal distribution $TN(\lambda_i|\boldsymbol{\alpha})$, with right truncation for $Z_i = 0$, and left truncation for $Z_i = 1$.

A Gibbs sampling procedure for the LNM model has been implemented in the DMU-package (Madsen and Jensen, 2004).

Data. A data set consisting of monthly SCC records collected between 10 and 315 days post partum from 10,000 1st lactation Danish Holstein (**DHF**) cows was extracted from the Danish national cattle database. To be included, the following criteria should be fulfilled: Calving between 1990 and 2003, age at calving between 18 and 38 months, and belong to a herd with at least 5 primiparous cows per year in the period 1999 – 2003. Summary statistics for the sampled data are given in Table 1.

Table 1. Summary statistics of the data set used in the statistical analysis

Records no.	84,372
Cows no.	10,000
Records per cow, (mean)	8.4
Animals in pedigree, no.	19,778
Herds, no	374
Herd-test-day classes, no.	11,325
SCS[1], mean (SD)	4.56 (1.15)

[1]SCS = ln (SCC * 10^{-3})

Model. The following LNM model was fitted;

$$\begin{bmatrix}\mathbf{SCS}\\ \boldsymbol{\lambda}\end{bmatrix} = \begin{bmatrix}\mathbf{X}\boldsymbol{\beta}_{0_{SCS}}+\mathbf{M}_z\boldsymbol{\beta}_{1_{SCS}}+\mathbf{W}_h\mathbf{h}_{SCS}+\mathbf{W}_a\mathbf{a}_{0_{SCS}}+\mathbf{M}_z\mathbf{W}_a\mathbf{a}_{1_{SCS}}+\mathbf{W}_p\mathbf{p}_{0_{SCS}}+\mathbf{M}_z\mathbf{W}_p\mathbf{p}_{1_{SCS}}+\mathbf{e}_{SCS}\\ \mathbf{X}\boldsymbol{\beta}_\lambda+\mathbf{W}_h\mathbf{h}_\lambda+\mathbf{W}_a\mathbf{a}_\lambda+\mathbf{W}_p\mathbf{p}_\lambda+\mathbf{e}_\lambda\end{bmatrix}$$

where the vectors $\boldsymbol{\beta}_{0_{SCS}}$ and $\boldsymbol{\beta}_\lambda$ vectors include systematic effects of age at calving (in month classes); regression coefficients for days carrying calf, general heterosis, specific Danish **DHF** × American Holstein (**HF**) heterosis, gene proportions from DHF, HF and other breeds, Legendre polynomials of days in milk (**DIM**) up to 2nd order, and a Wilmink term $(e^{-0.09DIM})$, with **X** as the appropriate incidence matrix. The Wilmink term was included for modeling the rapid decrease in SCS in the beginning of the lactation, and the

factor -0.09 was adapted from the Finnish test-day model for SCS (Negussie, pers. comm. 2004) The vector $\boldsymbol{\beta}_{1_{SCS}}$ includes the systematic effect of mastitis on SCS level only, with an incidence matrix $\mathbf{M}_z$. The latter is a diagonalization of $\mathbf{Z}$, the vector containing the assumed health statuses for all observations.

The first 10,000 samples were discarded as burn-in. For the next 75.000 samples every 10th were stored. Convergence of the Gibbs chain was checked using the method of batching and visual inspection of trace plots.

Results and Discussion

Posterior means for the frequency of putative mastitis was .283 (SD .003) on a test-day basis. The overall mean estimated for test-day SCS of healthy cows was 4.4, while the corresponding level for cows with putative mastitis was 5.0. Hence, the "typical" SCC level for an IMI- cow was 80,000 cells/ml, vs. 151,000 cells/ml for an IMI+ cow.

Table 2. Posterior means and standard deviations (in brackets) of genetic, permanent environmental, residual and herd-test-day (co)variance components[1], and corresponding heritabilities and repeatabilities for test-day SCS and liability to putative mastitis.

	Trait	SCS (IMI–)	SCS (IMI+)	Liability
Genetic	SCS(IMI–)	.24 (.02)	.55 (.07)	.07 (.09)
	SCS (IMI+)	.16 (.02)	.35 (.05)	-.04 (.11)
	Liability	.02 (.02)	-.01 (.04)	.29 (.05)
Perma-nent Env.	SCS (IMI–)	.31 (.02)	.61 (.05)	.13 (.04)
	SCS (IMI+)	.22 (.02)	.42 (.05)	.25 (.06)
	Liability	.07 (.02)	.15 (.04)	.86 (.06)
Resi-dual	SCS (IMI–)	.13 (<.01)	0	0
	SCS (IMI+)	0	1.47 (.02)	0
	Liability	0	0	1
Herd-test-day	SCS		.07 (<.01)	.33 (.03)
	Liability		.05 (.01)	.37 (.03)
Herita-bility		.35 (.02)	.16 (.02)	.14 (.02)
Repea-tability		.81 (<.01)	.34 (.01)	.53 (.02)

[1]Correlations (in bold), variances (in italic) and covariances

Posterior means for (co)variance components for SCS (IMI-), SCS (IMI+) and liability to putative mastitis are given in Table 2. Generally, as expected, a higher degree of SCS variation was found for IMI+ compared with IMI-. By going from IMI- to IMI+, genetic and PE variance increased by 30-50%, while the residual variance increased by more than 1000%. Hence, SCS heritability and repeatability were considerably lower for IMI+ (.15 and .34, resp.) than for IMI- (.35 and .81, resp.). Analyzing the same data using a standard linear repeatability test-day model, heritability and repeatability estimates of .12 and .45 were obtained.

The most striking result, however, was the high heritability estimate obtained for liability to putative mastitis (.14), which is twice the estimates obtained for liability to veterinary-treated clinical mastitis in cross-sectional studies (Heringstad *et al.,* 2001) and substantially higher than most estimates obtained from longitudinal analysis of clinical mastitis in specific time-periods (Heringstad *et al.,* 2003). It should be noticed that putative mastitis based on SCC records is probably more closely connected with subclinical mastitis rather than the clinical cases, due to the fact that the SCC records only covers the test-days on which samples are actually taken, which often excludes test-days on which clinical mastitis is detected and treated. Further, clinical mastitis is only recorded when actual treatment is started, which is a decision made by the farmer and/or the veterinarian and therefore involves some subjective factors, while SCC is routinely recorded and objectively measured without any human decision-making interfering with the result of the analysis. These factors may contribute the higher heritability estimate found for putative mastitis using a mixture model. Estimated genetic and permanent environmental correlations between SCS (IMI-) and SCS (IMI+) were .55 and .61, indicating that cows having a high SCS when healthy are also more likely to have a higher SCS when having an IMI. The estimated genetic correlations between SCS and liability to putative mastitis were .07 and -.04 for SCS (IMI-) and SCS (IMI+). This indicated that SCS level in both healthy and diseased animals have rather weak genetic relationships with putative mastitis resistance. Hence, the results of this study does not support the hypothesis that selection for lower baseline SCS would result in deterioration of mastitis resistance. However, cows having positive permanent environmental effects for SCS level (particularly when diseased) seem to be slightly more susceptible

to IMI. A similar, but more expressed, positive correlation was estimated between herd-test-day effects for SCS (irrespective of IMI status) and liability to putative mastitis.

Conclusions

Mixture models are useful for identifying hidden structures affecting data, such as unrecorded cases of mastitis affecting test-day SCS. Using a liability normal mixture model, variance components for SCS from healthy udders were lower and repeatability and heritability for SCS higher, compared with SCS from infected udders. Further, cows having high SCS when healthy seem more likely to have high SCS also when infected. Genetic correlations between SCS and liability to putative mastitis were close to zero, irrespective of IMI status, while environmental correlations (PE and herd-test-day) were moderately positive. The heritability for liability to putative mastitis on a test-day level was substantially higher than most commonly reported estimates for liability to clinical mastitis. Hence, it seems promising to base selection on EBV for liability to putative mastitis rather than crudely selecting for lower SCS level. Still, no other data sources besides SCS test-day data are needed. The estimates obtained in this study are based on a limited data set and should therefore be verified in future studies with more data. Further, the genetic correlation between liability to putative mastitis (based on the LNM model) and liability to clinical mastitis need to be estimated in order to assess the usefulness of the new methodology in genetic selection.

References

Detilleux, J. & Leroy, P.L. 2000. *J. Dairy Sci. 83(10),* 2341-2349.

Gianola, D., Ødegård, J., Heringstad, B., Klemetsdal, G., Sorensen, D., Madsen, P., Jensen, J. & Detilleux, J. 2004. *Genet. Sel Evol. 36,* 3-27.

Heringstad, B., Rekaya, R., Gianola, D., Klemetsdal, G. & Weigel, K.A. 2001. *J. Dairy Sci. 84,* 2337-2346.

Heringstad, B., Chang, Y.M., Gianola, D. & Klemetsdal, G. 2003. *J. Dairy Sci. 86,* 2676-2683.

Madsen, P. & Jensen, J. 2004. *A user's guide to DMU.* Version 6, release 4.5, DJF, Foulum, Denmark. 31pp.

Mclachlan, G. & Peel, D. 2000. *Finite mixture models.* Wiley, New York.

Ødegård, J., Jensen, J., Madsen, P., Gianola, D., Klemetsdal, G. & Heringstad, B. 2003. *J. Dairy Sci. 86,* 3694-3703.

Ødegård, J., Madsen, P., Gianola, D., Klemetsdal, G., Jensen, J., Heringstad, B. & Korsgaard, I.R. 2005. *J. Dairy Sci. 88,* 2652-2659.