Recursive relationships between milk yield and somatic cell score of Canadian Holsteins from finite mixture random regression models

J. Jamrozik¹ and L. R. Schaeffer

Centre for Genetic Improvement of Livestock, Department of Animal and Poultry Science, University of Guelph, Guelph, Ontario, Canada N1G 2W1

ABSTRACT

Finite mixture, multiple-trait, random regression animal models with recursive links between phenotypes for milk yield and somatic cell score (SCS) on the same test-day were applied to first lactation Canadian Holstein data. All models included fixed herd-test-day effects and fixed regressions within region–age at calving–season of calving classes, and animal additive genetic and permanent environmental regressions with random coefficients. Causal links between phenotypes for milk yield and SCS were fitted separately for records from healthy cows and cows with a putative, subclinical form of mastitis. Bayesian methods via Gibbs sampling were used for the estimation of model parameters. Bayes factors indicated superiority of the model with recursive link from milk to SCS over the reciprocal recursive model and the standard multiple-trait model. Differences between models measured by other, single-trait model comparison criteria (i.e., weighted mean squared error, squared bias, and correlation between observed and expected data) were negligible. Approximately 20% of test-day records were classified as originating from cows with mastitis in recursive mixture models. The proportion of records from cows infected with mastitis was largest at the beginning of lactation. Recursive mixture models exhibited different distributions of data from healthy and infected cows in different parts of lactation. A negative effect of milk to SCS (up to -0.15 score points for every kilogram of milk for healthy cows from 5 to 45 d in milk) was estimated for both mixture components (healthy and infected) in all stages of lactation for the most plausible model. The magnitude of this effect was stronger for healthy cows than for cows infected with mastitis. Different patterns of genetic and environmental correlations between milk and SCS for healthy and infected records were revealed, due to heterogeneity of structural coefficients between mixture components. Estimated breeding values for SCS from

the best fitting model for sires of infected daughters were more related to estimated breeding values for the same trait from the regular multiple-trait model than evaluations for sires of mastitis-free cows.

Key words: causal relationship, milk yield, mixture model, somatic cell score

INTRODUCTION

Phenotypic correlation between milk yield and SCS in dairy cattle may be generated by causal relationships, and several different mechanisms driving these relationships can be hypothesized. High level of SCS associated with mammary infection can result in a decrease of milk production (infection effect), whereas high-producing cows are usually more prone to mastitis and thus they can express elevated SCS (stress effect). On the other side, concentration of somatic cells in milk may depend on the overall level of milk yield, resulting in lower levels of SCS for higher yielding cows with the same overall cell count (dilution effect). Finally, a certain level of SCS might be necessary to protect a cow against infections, and very low SCS may result in decrease of milk yield due to suboptimal immunological protection (fitness effect). All these effects could operate as recursive (trait A affects trait B and the reverse does not occur) or simultaneous (trait A affects trait B and, at the same time, trait B affects trait A) links in the framework of structural equation models (SEqM; Gianola and Sorensen, 2004). These models allow the modeling of causal relationships between phenotypes in addition to fitting genetic and environmental correlations. Milk yield and SCS in dairy cows have been analyzed by SEqM in several studies. All published research indicated a negative effect of SCS on milk yield. Thus, the negative phenotypic relationship between milk and SCS could be more likely explained by an infection than by a dilution effect. de los Campos et al. (2006) investigated relationships between SCS and milk yield in the first lactation of Norwegian Red cows. Model comparison statistics favored the model with a negative effect of SCS on milk yield over the 3 other models fitted (i.e., absence of recursive effects; an effect of milk

Received May 25, 2010.

Accepted July 26, 2010.

¹Corresponding author: jjamrozik@uoguelph.ca

yield on SCS; and simultaneous effects between milk yield and SCS). Wu et al. (2007) inferred relationships between first-lactation SCS and milk yield in the same population. Results suggested a large negative effect of SCS on milk yield and a small, positive reciprocal effect. Negative phenotypic effects of SCS on milk yield and a smaller positive reciprocal effect were also estimated in Canadian Holsteins using random regression models with simultaneous causal links between these traits on the same test-day (**TD**) (Jamrozik et al., 2010). Less plausible recursive parameterizations indicated a negative effect from SCS to milk yield (infection effect) or a negative effect from milk to SCS (dilution effect). The same authors indicated heterogeneity of phenotypic relationships between milk and SCS in different lactations and in different parts of a specific lactation. Heritability estimates from SEqM in these studies were similar to those from the standard mixed model, but some correlations differed considerably among models. Rankings of animals with respect to EBV for milk yield and SCS did not differ between multiple-trait (**MT**) models and SEqM (Jamrozik et al., 2010).

Relationships between milk yield and SCS may depend on the health status of the animal. In particular, the pattern of relationships between milk yield and SCS in cows infected with mastitis could be different from those that are mastitis free. This refers to the direction of the causal link as well as the strength of associations. The information on cow health, however, is seldom available in population genetic studies. Finite mixture models have been proposed for analysis of SCS in the absence of information regarding infection status of the cow. These models assume possible different distributions of SCS for healthy and infectious cows to reflect the fact that SCS might be different traits for these 2 groups (Dettelleux and Leroy, 2000). Mixture models have been applied to analysis of SCS of US Holsteins (Boettcher et al., 2007) and Danish Holsteins (Madsen et al., 2008). Jamrozik and Schaeffer (2010) used a finite mixture, random regression model for a joint 2-trait analysis of TD milk yield and SCS of Canadian Holsteins. Results of this study indicated that approximately 20% of TD records originated from cows with a putative, subclinical form of mastitis. The proportion of records from cows with putative mastitis was the largest at the beginning of lactation. The mixture model exhibited different distributions of data from healthy and infected cows in different parts of lactation: records from sick cows were characterized by larger (smaller) means for SCS (milk) and larger variances. Residual and daily genetic and environmental correlations between milk and SCS were smaller from the mixture model compared with the MT estimates. Heritabilities of both traits differed significantly among

health status from the mixture model and from the MT model estimates.

Exploring causal relationships between milk and SCS in the framework of finite mixture models, as proposed by Wu et al. (2010), might give new insight in the analysis of these traits. Our objectives were therefore to apply a 2-trait, finite mixture, random regression model with recursive links between phenotypes for a joint analysis of milk yield and SCS on the same TD of first-lactation Canadian Holsteins. Co-variance components, genetic parameters, and sire EBV for milk and SCS were compared with estimates from the standard MT model.

MATERIALS AND METHODS

Data

First-lactation TD records of Canadian Holstein cows, with calving years from 1988 to 2007, were obtained from the Canadian Dairy Network (Guelph, ON, Canada) as a subset of the national genetic evaluation data. A small, computationally manageable data set was selected by random sampling of herds with a minimum of 50 cows each and with a minimum of 7 TD records per cow. This resulted in 214,813 TD milk (kg) and SCS records on 25,950 cows. Each cow was required to have both milk yield and SCS recorded on a given TD. Somatic cell scores were calculated from SCC as follows: $SCS = \log_2(SCC/100,000) + 3$. Edits on DIM excluded records before 5 and after 305 d. The number of herd-test-day (**HTD**) classes was 23,984, and the number of region-age-season of calving (**RAS**) levels was 80; pedigree data included 66,244 animals. The data included TD records from healthy cows and cows with a subclinical form of mastitis. Data from cows with clinical signs of mammary infection are not typically stored for genetic evaluation purposes. The same data set was used by Jamrozik and Schaeffer (2010) for fitting a regular mixture model.

Models

The finite mixture model assigns records on milk and SCS on a given TD into 1 of the 2 components, depending on the unknown health status of the cow. Days in milk in lactation were split into 4 disjoint intervals: 5–45, 46–115, 116–265, and 266–305 DIM, as in Schaeffer et al. (2000), and group membership (healthy or infected) of the j th TD record in the k th DIM interval on the i th cow was defined by a binary vector $\mathbf{q}_k$, where $q_{ijk} = 0$ indicated records from a healthy (**IM**–) cow and $q_{ijk} = 1$ corresponded to records from an infected (**IM**+) cow.

The equation for the mixture model with recursive links describing milk and SCS for the i th cow on the j th TD in the k th interval of lactation, in the r th class of HTD effect, for the s th level of RAS effect, and for the m th mixture component can be written following Ødegård et al. (2003) and Wu et al. (2007):

$$\begin{aligned} \Lambda_{ijk} \mathbf{y}_{ijkms} = & \mathbf{X}_{fij} \beta_{fs} + \mathbf{X}_{hij} \beta_{hr} + q_{ijk} \mathbf{X}_{dij} \beta_{dk} \\ & + \mathbf{Z}_{ij} \mathbf{a}_i + \mathbf{W}_{ij} \mathbf{p}_i + \mathbf{e}_{ijkms}, \end{aligned} \quad [1]$$

where $\mathbf{y}_{ijkms}$ = vector (2×1) of observations (milk and SCS) on the i th cow for the j th TD; β_{fs} = vector of fixed regression coefficients for the s th RAS class; β_{hr} = vector (2×1) of fixed effects of the r th HTD class; β_{dk} = vector (2×1) of fixed effect corresponding to data from infected cows in the k th DIM interval; $\mathbf{a}_i$ = vector of random animal additive genetic effects for the i th animal; $\mathbf{p}_i$ = vector of random permanent environmental (PE) effects; $\mathbf{e}_{ijkms}$ = vector of random residuals, $\mathbf{X}_{fij}$, $\mathbf{X}_{hij}$, $\mathbf{X}_{dij}$, $\mathbf{Z}_{ij}$, $\mathbf{W}_{ij}$ = known incidence matrices, and $\Lambda_{ijk} = (1 - q_{ijk}) \Lambda^{(k0)} + q_{ijk} \Lambda^{(k1)}$, where

$$\Lambda^{(km)} = \begin{bmatrix} 1 & -\lambda_{M \leftarrow SCS}^{(km)} \\ -\lambda_{M \rightarrow SCS}^{(km)} & 1 \end{bmatrix},$$

which was the matrix of structural coefficient parameters (effect of change caused by the phenotype of the other trait) between phenotypes for milk yield and SCS in the k th DIM interval for the m th mixture component ($m = 0$ or 1). The elements of β_{dk} included mean differences between “sick” and “healthy” components for milk and SCS in the k th DIM interval. Additive genetic and PE effects were regressions on DIM, and all regressions (fixed and random) were modeled with orthogonal Legendre polynomials of order 4. Assume that elements of $\mathbf{a}_i$ and $\mathbf{p}_i$ are ordered as traits within regression coefficients.

Parameters of model [1] are called system parameters. Proper independent prior normal distributions [$\sim N(\mathbf{0}, \mathbf{I}10,000)$] were assumed for all fixed effects of model [1]. Random effects were also a priori normally distributed, conditionally on genetic ($\mathbf{G}$) and PE ($\mathbf{P}$) system covariance components; that is, $\mathbf{a} | \mathbf{G} \sim N(\mathbf{0}, \mathbf{A} \otimes \mathbf{G})$, $\mathbf{p} | \mathbf{P} \sim N(\mathbf{0}, \mathbf{I} \otimes \mathbf{P})$, where prior distributions for $\mathbf{G}$ and $\mathbf{P}$ were 10-dimensional (corresponding to 2 traits and 5 regression coefficients), independent inverted Wishart distributions with minimal numbers of degrees of freedom, and $\mathbf{I}$ and $\mathbf{A}$ were identity and additive relationship matrices, respectively. System residual co-variance matrices ($\mathbf{R}_{km}$) were assumed different among mixture components and DIM intervals. To achieve identification of parameters in recursive models, system residual

co-variances were all set to zero. Prior distributions for system residual variances for milk and SCS were therefore independent inverted chi-squared distributions, also with minimal numbers of degrees of freedom. Elements of $\mathbf{q}_k$ were assumed to be independent and following the same Bernoulli distribution, $\text{Br}(Q_k)$, where Q_k was the proportion of IM+ records in the k th DIM interval (stage of lactation specific mixing proportion). The prior distributions for Q_k were beta distributions with both hyper-parameters equal to 2.

When Λ_{ijk} is a matrix of full rank and with a complete loading of factors (each factor affects all traits), then equation [1] can be rewritten (Wu et al., 2010) as

$$\begin{aligned} \mathbf{y}_{ijkms} = & \Lambda_{ijk}^{-1} \mathbf{X}_{fij} \beta_{fs} + \Lambda_{ijk}^{-1} \mathbf{X}_{hij} \beta_{hr} \\ & + q_{ijk} \Lambda_{ijk}^{-1} \mathbf{X}_{dij} \beta_{dk} + \Lambda_{ijk}^{-1} \mathbf{Z}_{ij} \mathbf{a}_i + \Lambda_{ijk}^{-1} \mathbf{W}_{ij} \mathbf{p}_i \\ & + \Lambda_{ijk}^{-1} \mathbf{e}_{ijkms} = \mathbf{X}_{fij} (\mathbf{I} \otimes \Lambda_{ijk}^{-1}) \beta_{fs} + \mathbf{X}_{hij} \Lambda_{ijk}^{-1} \beta_{hr} \\ & + q_{ijk} \mathbf{X}_{dij} \Lambda_{ijk}^{-1} \beta_{dk} + \mathbf{Z}_{ij} (\mathbf{I} \otimes \Lambda_{ijk}^{-1}) \mathbf{a}_i + \mathbf{W}_{ij} \\ & (\mathbf{I} \otimes \Lambda_{ijk}^{-1}) \mathbf{p}_i + \Lambda_{ijk}^{-1} \mathbf{e}_{ijkms} = \mathbf{X}_{fij} \beta_{fskm}^* \\ & + \mathbf{X}_{hij} \beta_{hrkm}^* + q_{ijk} \mathbf{X}_{dij} \beta_{dk}^* + \mathbf{Z}_{ij} \mathbf{a}_{ikm}^* \\ & + \mathbf{W}_{ij} \mathbf{p}_{ikm}^* + \mathbf{e}_{ijkms}^*, \end{aligned}$$

where $\mathbf{I}$ is an identity matrix of order equal to number of regression coefficients (5).

The conditional distribution of $\mathbf{y}_{ijkms}$ was

$$\begin{aligned} \mathbf{y}_{ijkms} | \beta_f, \beta_h, \beta_d, \mathbf{a}, \mathbf{p}, \mathbf{R}_{km}, q_{ijk}, \Lambda_{ijk} \sim & N[\mathbf{X}_{fij} \beta_{fskm}^* \\ & + \mathbf{X}_{hij} \beta_{hrkm}^* + q_{ijk} \mathbf{X}_{dij} \beta_{dk}^* + \mathbf{Z}_{ij} \mathbf{a}_{ikm}^* \\ & + \mathbf{W}_{ij} \mathbf{p}_{ikm}^*, \mathbf{R}_{ijkm}^*], \end{aligned}$$

where $\mathbf{R}_{ijkm}^* = (1 - q_{ijk}) \mathbf{R}_{k0}^* + q_{ijk} \mathbf{R}_{k1}^*$; $\mathbf{R}_{km}^* = \Lambda^{(km)-1} \mathbf{R}_{km} \Lambda^{(km)-1}$, ($m = 0$ or 1) are 2×2 residual covariance matrices in the k th DIM interval for IM- and IM+ records, respectively. Observations on different cows and TD were conditionally independent. System $\mathbf{G}$ and $\mathbf{P}$ covariance matrices generate a heterogeneous (depending on the mixture component and DIM interval in lactation) population covariance structure, namely

$$\text{var}(\mathbf{a}_{ikm}^*) = \mathbf{G}_{km}^* = (\mathbf{I} \otimes \Lambda^{(km)-1}) \mathbf{G} (\mathbf{I} \otimes \Lambda^{(km)-1}), \quad [2]$$

$$\text{var}(\mathbf{p}_{ikm}^*) = \mathbf{P}_{km}^* = (\mathbf{I} \otimes \Lambda^{(km)-1}) \mathbf{P} (\mathbf{I} \otimes \Lambda^{(km)-1}). \quad [3]$$

Genetic parameters (i.e., heritabilities, genetic and phenotypic correlations) for traits can then be calculated in a usual manner from elements $\mathbf{G}_{km}^*$, $\mathbf{P}_{km}^*$, and $\mathbf{R}_{km}^*$. Similarly, breeding values for animal i ($\mathbf{a}_{ikm}^*$) can be derived from the system parameter ($\mathbf{a}_i$) as

$$\mathbf{a}_{ikm}^* = (\mathbf{I} \otimes \mathbf{\Lambda}^{(km)-1}) \mathbf{a}_i, \quad [4]$$

and mean differences (on the population level) between records from infected and healthy cows can be calculated as

$$\beta_{dk}^* = \mathbf{\Lambda}^{(k1)-1} \beta_{dk}. \quad [5]$$

Again, structural coefficients for different mixture components and DIM interval in lactation (km) may generate heterogeneous breeding values with respect to the health status of an animal and stage of lactation.

Two specific recursive mixture models were fitted, namely:

$\mathbf{M} \leftarrow \mathbf{SCS}$ = model with causal links from SCS to milk yield on the same TD [$\lambda^{(km)}_{\mathbf{M} \rightarrow \mathbf{SCS}} = 0$] and

$\mathbf{M} \rightarrow \mathbf{SCS}$ = model with causal links from milk to SCS on the same TD [$\lambda^{(km)}_{\mathbf{M} \leftarrow \mathbf{SCS}} = 0$].

The standard MT model for milk and SCS was the same as the recursive mixture models assuming all elements of $\mathbf{q}_k$ and all structural coefficients equal to zero, and with unstructured residual covariance matrices.

Methods

Gibbs sampling scheme for the scalar 2-component mixture model (Ødegård et al., 2003) was generalized for the multiple-trait (2 traits) scenario and a heterogeneous data structure, and then combined with the algorithm designed for the SEqM (Wu et al., 2007). The system mean difference between IM– and IM+ groups (elements of β_{dk}) was required to be larger (smaller) than 0 for SCS (milk yield). Conditional distributions for the recursive mixture models were multivariate normal for the system location parameters, univariate normal for structural coefficients, inverted Wishart for the system genetic and PE co-variance components, independent inverted chi-squared distribution for the system residual variances, Bernoulli for the group memberships, and beta distributions for the mixing proportions. The sampling for the MT model followed standard algorithms. Single chains of 120,000 samples were generated for all models. Convergence of Gibbs chains was determined by visual inspection of trace plots for selected parameters. All posterior inferences were based on 100,000 samples of the population parameters (derived from the system parameters through appropriate transformations of [2] to [5]) after burn-in.

Models were compared via Bayes factors (Kass and Raftery, 1995). The marginal likelihood was estimated by the harmonic mean of the likelihood values from the Gibbs chain as $p(\mathbf{y} \mid \mathbf{M} = \mathbf{M}_t) = \{1/m \sum [p(\mathbf{y} \mid \boldsymbol{\theta}_t^i, \mathbf{M}_t)]^{-1}\}^{-1}$, where $\boldsymbol{\theta}_t^i$ ($i = 1, 2, \dots, m$) were draws obtained from the posterior distribution under model t with parameters $\boldsymbol{\theta}_t$. In addition, goodness of fit of the models were compared by

weighted (by a residual variance) mean squared error

$$(\mathbf{WMSE}) = 1/n \sum \sum [y_{ik} - E(y_{ik} \mid \boldsymbol{\theta})]^2 / \sigma_{ek}^2,$$

and percentage of squared bias ($\mathbf{PSB}$) = $\sum [y_i - E(y_i \mid \boldsymbol{\theta})]^2 / \sum y_i^2 \times 100$, defined as in Ali and Schaeffer (1987), and the Pearson correlation between observations and their expected values:

$$\mathbf{R} = \text{Corr}[\mathbf{y}, E(\mathbf{y} \mid \boldsymbol{\theta})].$$

All 3 statistics were calculated separately for both traits; n was the total number of observations, $\boldsymbol{\theta}$ and σ_{ek}^2 were posterior means for the model parameters, and residual variance in the k th DIM interval, respectively.

Test-day records were further assigned a posteriori to healthy (**M50–**) and sick (**M50+**) categories, based on the 50% threshold and the posterior mean of the group membership binary parameter (q_{ijk}) for each observation, following Jamrozik and Schaeffer (2010). Records for which posterior average of q_{ijk} was smaller (larger) than 0.5 were assigned to the M50– (M50+) groups. Posterior probability of mastitis (**PPM**) for each TD record from the recursive mixture models was estimated following Ødegård et al. (2003). Finally, posterior means and standard deviations of co-variance components and genetic parameters on a daily basis, and EBV for cumulative milk yield and average daily SCS within 4 DIM intervals (for sires with at least 200 TD records) from the recursive mixture and the MT models were compared. Heterogeneous matrices of structural coefficients generated heterogeneous co-variance matrices ($\mathbf{G}_{km}^*$, $\mathbf{P}_{km}^*$, and $\mathbf{R}_{km}^*$) and heterogeneous location parameters of the model. Thus, potentially different genetic, PE, and residual variances (and respective genetic parameters) as well as EBV would exist, depending on the health status of the animal (IM– or IM+) and DIM within lactation (4 intervals).

RESULTS

Model Comparison

Model comparison criteria (Table 1) showed superiority of both recursive mixture models for milk yield and

Table 1. Model comparison statistics¹

Model ²	Log-likelihood	WMSE		PSB (%)		R	
		Milk	SCS	Milk	SCS	Milk	SCS
MT	−746,450	0.61	0.63	0.39	5.85	0.97	0.91
M ← SCS	−647,317	0.62	0.63	0.38	5.85	0.96	0.91
M → SCS	−633,046	0.61	0.64	0.39	5.84	0.96	0.90

¹Log-likelihood = marginal log-likelihood; WMSE = weighted mean squared error; PSB = percentage of squared bias; R = correlation between observed and expected observations.

²MT = multiple-trait model; M ← SCS = mixture model with recursive link from SCS to milk; M → SCS mixture model with recursive link from milk to SCS.

SCS over the standard MT model. Bayes factors indicated that the model with recursive link from milk to SCS (M → SCS) was the most plausible specification. Differences between MT and recursive mixture models for WMSE, PSB, and correlation between observed and expected data were negligible; both recursive models gave practically the same values for all these criterions. Slightly better fit for milk yield data than for SCS was achieved in terms of WMSE; PSB indicated substantially better goodness of fit for TD milk records in all models. Observations and their expected values ranked more similar for milk than for SCS, indicating again better accuracy of fitting milk records. Marginal log-likelihood for the standard mixture model (no recursive links) fitted to the same data in Jamrozik and Schaeffer (2010) was equal to −645,855. This would suggest that this model more likely generated the data than the recursive M ← SCS parameterization. Estimates of the remaining model comparison statistics for the regular mixture model (results not shown) were similar to those for the recursive mixture models.

Mixture and Structural Parameters

Estimates of mixture and structural parameters for the recursive mixture models are in Table 2. The model

with recursive link from milk to SCS (M → SCS) resulted in slightly larger proportions of TD records in IM− compared with the M ← SCS model (0.28 vs. 0.26 for the 5–45 DIM interval). Frequency of TD records from cows with putative subclinical mastitis was the largest at the beginning of lactation and decreased slightly with DIM in lactation. Frequencies of TD records in the M50+ category were lower than corresponding mixing frequencies across both models and all parts of lactation. Again, the M → SCS model gave slightly larger estimates of these parameters compared with the M ← SCS model. Plots of average milk yield and SCS by DIM for M50− and M50+ categories are in Figure 1. Differences between lactation curves were small for milk yield, with slightly larger averages for M50− records at the beginning of lactation. Curves for SCS, however, were distinctly different. The level of SCS in the M50+ category was about 2 score points higher than corresponding averages for the M50− group.

Milk yield records from infected cows were characterized by smaller means (parameters β^*_{dk}) than TD records from healthy cows; the opposite was observed for SCS. The shift in averages for a given trait was larger for the model in which this trait served as a covariate in the model for the other trait. Differences in averages for

Table 2. Estimates (posterior SD in parentheses) of mixing proportion (Q), frequency of 50% mastitic records (M-50), difference (β^*_{d4}) between sick (IM+) and healthy (IM−) groups for milk (M) and SCS, and structural coefficients (λ) from recursive mixture models, by DIM interval

Model ¹	Parameter	DIM interval			
		5–45	46–115	116–265	266–305
M ← SCS	Q	0.26 (0.01)	0.19 (0.01)	0.16 (0.01)	0.17 (0.01)
	M-50	0.19	0.13	0.11	0.11
	β^*_{dM}	−0.6 (0.09)	−0.3 (0.05)	−0.2 (0.04)	−0.5 (0.11)
	β^*_{dSCS}	1.4 (0.04)	1.2 (0.03)	1.2 (0.02)	1.1 (0.04)
	λ for IM−	0.26 (0.03)	0.03 (0.02)	0.14 (0.01)	0.26 (0.03)
	λ for IM+	−0.26 (0.03)	−0.19 (0.02)	−0.13 (0.01)	−0.08 (0.03)
M → SCS	Q	0.28 (0.01)	0.21 (0.01)	0.17 (0.01)	0.18 (0.01)
	M-50	0.20	0.15	0.12	0.11
	β^*_{dM}	−1.1 (0.09)	−0.6 (0.06)	−0.5 (0.04)	−0.8 (0.09)
	β^*_{dSCS}	0.2 (0.03)	0.1 (0.02)	0.1 (0.02)	0.1 (0.04)
	λ for IM−	−0.15 (0.01)	−0.14 (0.01)	−0.14 (0.01)	−0.14 (0.01)
	λ for IM+	−0.10 (0.01)	−0.11 (0.01)	−0.10 (0.01)	−0.10 (0.01)

¹M ← SCS = mixture model with recursive link from SCS to milk; M → SCS mixture model with recursive link from milk to SCS.

Figure 1. Average milk yield and SCS from M50– [posterior mean of classification variable (q_{ijk}) < 0.5] and M50+ [posterior mean of classification variable (q_{ijk}) > 0.5] categories based on the mixture model with recursive link from SCS to milk yield, by DIM. Color version available in the online PDF.

milk yield showed, in general, a decreased trend with DIM (with the exception for 266–305 DIM); estimates for SCS seemed to be more stable across lactation.

Estimates of structural coefficients differed significantly between IM– and IM+ groups. Model $M \rightarrow SCS$ (the most plausible specification) indicated strong dilution effects (a decrease in SCS concentration with larger milk yields) for both mixture components. The magnitude of structural coefficients from milk to SCS was almost 50% larger for healthy cows (IM–) compared with infected cows, with a relatively stable trend across lactation. Model $M \leftarrow SCS$ indicated the possibility of a fitness effect from SCS to milk for records from healthy cows; that is, an increase in TD milk yield with higher levels of SCS. The pattern of changes in structural coefficient from SCS to milk with DIM followed a typical lactation curve shape for SCS, with the largest values at both ends of lactation. The opposite relationship between milk and SCS (a decrease of milk yield with larger SCS) was observed for records from infected cows for the same model, indicating a possibility of an infection effect for this mixture component. Structural coefficients from SCS to milk yield showed linear, decreasing in magnitude, trend with DIM for infected cows.

Average phenotypes for SCS corrected for milk yield using estimates of structural coefficients from the $M \rightarrow SCS$ model for both mixture components are in Figure 2. Milk-corrected SCS for both mixture components

had significantly higher level than original SCS; healthy cows showed a lower (1 score point on average) level of SCS than cows with putative mastitis. Both milk-corrected SCS curves were almost linear (opposite to a standard lactation curve for SCS), slightly decreasing with DIM, with a little more variation (most likely due to smaller sample sizes) for the IM+ component. Lactation curves for milk corrected for the SCS level are in Figure 3. Differences between curves were less profound than in those for SCS; infected cows gave slightly more SCS-corrected milk in the middle of lactation.

Co-variance Components and Genetic Parameters

Estimates of population residual co-variances from milk yield and SCS from the MT and both recursive mixture models are in Table 3. Variances for both traits showed a typical, decreasing trend with DIM. Records from healthy cows were characterized by significantly smaller residual variances than those from cows with subclinical mastitis, indicating different distribution for 2 mixture components. Differences in residual variances were small between the 2 recursive models. Larger differences between all 3 models and mixture components were observed for estimates of residual correlations. Estimates differed in signs between IM– (positive) and IM+ (negative) components for the $M \leftarrow SCS$ model. The model with a causal link from milk to SCS (the best model) generated relatively large, negative residual

Figure 2. Average overall and milk-corrected SCS from M50– [posterior mean of classification variable (q_{ijk}) < 0.5] and M50+ [posterior mean of classification variable (q_{ijk}) > 0.5] categories from the mixture model with recursive link from milk to SCS, by DIM. Color version available in the online PDF.

co-variability, especially for records from healthy cows. Estimates of residual co-variance components from the $M \leftarrow SCS$ model were more similar to parameters obtained from the regular mixture model (Jamrozik and Schaeffer, 2010).

Estimates of population daily genetic variances are in Figures 4 and 5 for milk and SCS, respectively. Genetic variances for recursive models and their mixture com-

ponents for milk yield were slightly larger compared with multiple-trait estimates. The opposite trend was noted for daily SCS. Differences between IM– and IM+ estimates for both traits and both mixture recursive models were less visible. Similar trends, with slightly larger differences between MT and recursive models, were observed for estimates of daily PE variances for both traits (results not shown).

Figure 3. Average overall and SCS-corrected milk yield from M50– [posterior mean of classification variable (q_{ijk}) < 0.5] and M50+ [posterior mean of classification variable (q_{ijk}) > 0.5] categories from the mixture model with recursive link from SCS to milk yield, by DIM. Color version available in the online PDF.

Figure 4. Daily genetic variance of milk yield from the multiple-trait model (MT), and mixture models with recursive link from SCS to milk ($M \leftarrow SCS$) and recursive link from milk to SCS ($M \rightarrow SCS$), separately for healthy (IM–) and sick (IM+) animals. Color version available in the online PDF.

Relatively large differences between residual variances for 2 mixture components resulted in significantly different estimates of heritabilities for records from healthy and infected cows, as shown in Figures 6 and 7 for milk yield and SCS, respectively. The largest heritabilities were for IM–, followed by MT estimates, and the IM+ category. Estimates from both recursive mixture models were almost the same as those from the regular mixture model (Jamrozik and Schaeffer, 2010).

Genetic correlations between milk yield and SCS on a daily basis (Figure 8) from the best model ($M \rightarrow SCS$) differed markedly between 2 mixture components: close to zero for IM– and about 0.25 for IM+, showing the opposite trend in magnitude to the corresponding residual correlations. Estimates for the remaining models and mixture components were in between these 2 extreme cases. For the $M \leftarrow SCS$ model, milk and SCS were less correlated in infected cows than in cows free of mastitis. Estimates of PE correlations between

Table 3. Estimates (posterior SD in parentheses) of population residual variance (σ^2) and residual correlation (r) for milk yield (M) and SCS from the multiple-trait and recursive mixture models for sick (IM+) and healthy (IM–) groups, by DIM interval

Model ¹	Mixture component	Parameter	DIM interval			
			5–45	46–115	116–265	266–305
MT		σ^2_M	7.7 (0.15)	4.8 (0.05)	3.9 (0.02)	3.4 (0.9)
		σ^2_{SCS}	1.1 (0.02)	0.9 (0.01)	0.7 (0.01)	0.6 (0.01)
		r	–0.2 (0.01)	–0.1 (0.01)	–0.1 (0.01)	–0.2 (0.02)
$M \leftarrow SCS$	IM–	σ^2_M	3.6 (0.14)	2.3 (0.04)	1.9 (0.02)	1.4 (0.07)
		σ^2_{SCS}	0.3 (0.01)	0.2 (0.01)	0.2 (0.01)	0.2 (0.01)
		r	0.07 (0.01)	0.01 (0.01)	0.05 (0.01)	0.09 (0.01)
	IM+	σ^2_M	19.6 (0.6)	15.6 (0.3)	14.7 (0.2)	14.4 (0.6)
		σ^2_{SCS}	3.3 (0.08)	3.2 (0.06)	2.8 (0.04)	2.5 (0.08)
		r	–0.11 (0.01)	–0.08 (0.01)	–0.06 (0.01)	–0.03 (0.01)
$M \rightarrow SCS$	IM–	σ^2_M	3.1 (0.12)	1.9 (0.04)	1.6 (0.02)	1.1 (0.06)
		σ^2_{SCS}	0.3 (0.01)	0.3 (0.01)	0.3 (0.01)	0.2 (0.01)
		r	–0.45 (0.01)	–0.36 (0.01)	–0.34 (0.01)	–0.36 (0.01)
	IM+	σ^2_M	21.1 (0.6)	15.9 (0.3)	14.8 (0.2)	14.5 (0.5)
		σ^2_{SCS}	3.3 (0.08)	3.4 (0.07)	2.9 (0.04)	2.7 (0.09)
		r	–0.26 (0.01)	–0.23 (0.01)	–0.23 (0.01)	–0.23 (0.01)

¹MT = multiple-trait model; $M \leftarrow SCS$ = mixture model with recursive link from SCS to milk; $M \rightarrow SCS$ mixture model with recursive link from milk to SCS.

Figure 5. Daily genetic variance of SCS from the multiple-trait model (MT), and mixture models with recursive link from SCS to milk ($M \leftarrow SCS$) and recursive link from milk to SCS ($M \rightarrow SCS$), separately for healthy (IM-) and sick (IM+) animals. Color version available in the online PDF.

milk and SCS were mostly negative (Figure 9). In contrast to genetic correlations, PE relationships between milk and SCS were the weakest (close to 0) for records from IM+ generated by the $M \rightarrow SCS$ model. Model $M \leftarrow SCS$ gave the strongest negative PE correlations

(around -0.3) for the IM+ group. All daily phenotypic correlations between milk and SCS were negative, and differences between models and mixture components were relatively small (Figure 10). For both recursive models, phenotypic correlations on the same TD for

Figure 6. Daily heritability of milk yield from the multiple-trait model (MT), and mixture models with recursive link from SCS to milk ($M \leftarrow SCS$) and recursive link from milk to SCS ($M \rightarrow SCS$), separately for healthy (IM-) and sick (IM+) animals. Color version available in the online PDF.

Figure 7. Daily heritability of SCS yield from the multiple-trait model (MT), and mixture models with recursive link from SCS to milk ($M \leftarrow SCS$) and recursive link from milk to SCS ($M \rightarrow SCS$), separately for healthy (IM-) and sick (IM+) animals. Color version available in the online PDF.

the healthy cows were slightly smaller than those corresponding to cows with putative mastitis. The extreme correlations (around -0.1) were estimates for the IM+ class from the $M \rightarrow SCS$ model.

EBV

Correlations between milk yield EBV for sires with at least 200 TD records from the MT and both recursive models were equal to 1.0 for all DIM intervals and both

Figure 8. Daily genetic correlations between milk yield and SCS from the multiple-trait model (MT), and mixture models with recursive link from SCS to milk ($M \leftarrow SCS$) and recursive link from milk to SCS ($M \rightarrow SCS$), separately for healthy (IM-) and sick (IM+) animals. Color version available in the online PDF.

Figure 9. Daily permanent environmental (PE) correlations between milk yield and SCS from the multiple-trait model (MT), and mixture models with recursive link from SCS to milk ($M \leftarrow SCS$) and recursive link from milk to SCS ($M \rightarrow SCS$), separately for healthy (IM-) and sick (IM+) animals. Color version available in the online PDF.

mixture components. The same values of correlations were estimated for SCS and MT versus $M \leftarrow SCS$ comparison. More differences in rankings of sires would be expected between the MT and $M \rightarrow SCS$ (the most plausible model) for SCS: correlations within 4 DIM intervals were 0.93, 0.86, 0.89, and 0.90 for the IM- component, and 0.98, 0.94, 0.95, and 0.95 for the IM+

category. Comparisons between $M \rightarrow SCS$ and $M \leftarrow SCS$ models gave all correlations >0.99 for both traits, all intervals in lactation, and within both mixture components.

Sires ranked almost identically for milk yield when 2 mixture components were compared within each recursive model. A similar trend was observed for SCS and

Figure 10. Daily phenotypic correlations between milk yield and SCS from the multiple-trait model (MT), and mixture models with recursive link from SCS to milk ($M \leftarrow SCS$) and recursive link from milk to SCS ($M \rightarrow SCS$), separately for healthy (IM-) and sick (IM+) animals. Color version available in the online PDF.

the $M \leftarrow SCS$ model. Slightly more rerankings would be expected between IM- and IM+ for SCS from the $M \rightarrow SCS$ model (correlations between EBV were 0.98 for all DIM intervals).

Correlations between recursive models for the posterior probability of mastitis were equal to 0.98 for all 4 DIM intervals. Correlations between PPM and EBV for SCS from the $M \rightarrow SCS$ model were significantly different from zero only in the 5–45 DIM interval (0.43 and 0.45 for IM- and IM+, respectively). Respectively, model $M \leftarrow SCS$ gave significantly different from zero correlations in the first 2 intervals: 0.45 for 5–45 DIM and 0.16 for 46–115 DIM, with the same values for both mixture components.

DISCUSSION

Finite mixture recursive models allowed separation of TD data for milk yield and SCS into 2 components: originating from cows with putative, subclinical mastitis and those free of mammary infection. Different recursive effects (from milk to SCS or from SCS to milk yield on the same TD) were simultaneously fitted for different mixture components. Mixture proportions for recursive mixture models were similar to those estimated by Jamrozik and Schaeffer (2010) for a regular mixture model. Mixing probabilities estimated in this study were larger than reported by Boettcher et al. (2007) for the US Holstein (5%) but lower than estimates of Madsen et al. (2008) for Danish Holsteins (30%). Discrepancies between different estimates could be caused by differences in populations and the models used for estimation.

Milk yield and SCS showed different distributions for IM- and IM+ records. Records from cows with mastitis were characterized by larger (smaller) means for SCS (milk) and substantially larger variances compared with performance of mastitis free cows. Results were in good agreement with estimates of Jamrozik and Schaeffer (2010) from a regular mixture model on the same data set.

Both direction and magnitude of causal relationships between phenotypes for milk and SCS on the same TD differed between mixture components. The most plausible model ($M \rightarrow SCS$) exhibited a negative phenotypic effect from milk to SCS (dilution effect), relatively homogeneous with respect to different stages of lactation. The magnitude of this effect was larger for records from healthy cows. This could be explained by a stress effect (an increase of SCS with larger milk yield) that would act in a different direction in cows from IM+ category. Larger milk yield for infected cows would also explain reduction of the magnitude of structural coefficient from milk to SCS in this mixture component. Overall

results, in terms of model plausibility, were somehow contrary to findings in other research. Wu et al. (2007) and Jamrozik et al. (2010) concluded that relationships between milk and SCS seemed to be driven by a combination of 2 mechanisms through simultaneous links. An infection effect caused a decrease in milk yield for cows with a higher SCS, and the stress effect resulted in an increase of SCS for higher producing cows. Results of de los Campos et al. (2006) also favored the model with a negative recursive effect from SCS to milk yield. Several possible explanations exist for these discrepancies. First, fitting separate causal links for different mixture components might reveal different mechanisms than obtained for the overall population. On the other hand, a dilution effect was revealed in other studies (i.e., Jamrozik et al., 2010) through a model with only a recursive link (no simultaneity). Mixture models with simultaneous links between milk and SCS [both $\lambda^{(km)}_{M \rightarrow SCS}$ and $\lambda^{(km)}_{M \leftarrow SCS}$ allowed to be different from zero] would require additional restrictions to identify parameters: one more restriction on residual co-variance component for each mixture and DIM combination to identify the likelihood, and exclusion restrictions for 1 set of fixed effects for each trait (Jamrozik et al., 2010). Attempts to fit a simultaneous mixture model for the current data revealed that additional exclusion restrictions on the effects for mean values for traits in the IM+ category would be necessary. This would make the simultaneous model too restricted (impractical).

Fitting separate causal links between milk and SCS allowed for heterogeneity of correlations between milk and SCS. Records from IM+ animals from the $M \rightarrow SCS$ model exhibited relatively larger genetic correlations than revealed for the IM- category. This would indicate that genetic associations between these 2 traits in healthy animals were weaker than in the group of infected cows. On the contrary, residual and PE correlations for the IM- group from the best plausible model were stronger than corresponding correlation for the IM+ category. Environmental association between milk yield and SCS on the same TD seemed therefore to be stronger for healthy cows.

Mixture recursive effects generated 2 EBV for each trait: one for IM- and one for IM+. Estimated breeding values for SCS from the $M \rightarrow SCS$ model were not perfectly correlated with EBV from the MT model. Ranking of sires was more similar between MT and IM+ than between the MT and IM- component, suggesting that EBV from the standard MT model would more correctly rank sires of infected cows for the purpose of selection against mastitis. Correlation between PPM and EBV for SCS from the regular mixture model was only 0.29 compared with 0.46 for the correlation between PPM and EBV for SCS from the multiple-trait

model (Jamrozik and Schaeffer, 2010). This indicated relatively poor performance of regular mixture model for generating EBV useful for selection against mastitis. Heterogeneity of EBV due to fitting different recursive links for different mixture components allowed for an increase in correlation between PPM and EBV for SCS in the first DIM interval to the level generated by the MT model (correlation of 0.45 for the IM+ group). This would make the recursive mixture model competitive (albeit still less practical) for the purpose of genetic selection of sires against mastitis. A liability-normal mixture model that would allow for different prior probabilities of putative mastitis for different TD records (Ødegård et al., 2005) would make a further improvement in modeling. Such a model would allow ranking animals based on the genetic component of posterior probability of mastitis.

Recursive models and mixture model parameterization for milk and SCS in dairy cows have recently been discussed in Jamrozik et al. (2010) and Jamrozik and Schaeffer (2010) in terms of usefulness, limitations, and possible generalizations. Fitting causal links between phenotypes in the framework of SEqM seems extremely attractive for revealing directional relationships between phenotypes. Utilizing mixture model methodology for revealing hidden structures in the data due to unknown subpopulation membership added an extra flavor to the structural equation approach. Causal models allow only for arguments in support of a given causal assumption. They are not meant, in general, to generate validated causal conclusions (Wu et al., 2010).

CONCLUSIONS

Finite mixture model with 2 components for TD milk and SCS exhibited different distributions of these traits because of the occurrence of putative mastitis in different parts of lactation. Negative phenotypic correlations between milk yield and SCS on the same TD in the first lactation of Canadian Holstein cows can more likely be explained by an effect of dilution than by other possible directional relationships between these traits. The negative effect from milk to SCS was stronger for records from healthy cows than for cows with putative subclinical mastitis. Heterogeneity of structural coefficients between mixture components resulted in different patterns of genetic and environmental correlations in healthy cows and cows infected with mastitis.

ACKNOWLEDGMENTS

The authors are grateful to DairyGen (Guelph, ON, Canada) and the Natural Science and Engineering Research Council (Ottawa, ON, Canada) for their financial support, and Canadian Dairy Network (Guelph, ON, Canada) for providing the data. We appreciate comments from 2 anonymous reviewers for improving the quality of the paper.

REFERENCES

- Ali, T. E., and L. R. Schaeffer. 1987. Accounting for covariances among test-day milk yields of dairy cows. *Can. J. Anim. Sci.* 67:637–644.
- Boettcher, P. J., D. Caraviello, and D. Gianola. 2007. Genetic analysis of somatic cell scores in US Holsteins with Bayesian mixture model. *J. Dairy Sci.* 90:435–443.
- de los Campos, G., D. Gianola, and B. Heringstad. 2006. A structural equation model for describing relationships between somatic cell score and milk yield in first-lactation dairy cows. *J. Dairy Sci.* 89:4445–4555.
- Detilleux, J., and P. J. Leroy. 2000. Application of a mixed normal mixture model for the estimation of mastitis-related parameters. *J. Dairy Sci.* 83:2341–2349.
- Gianola, D., and D. Sorensen. 2004. Quantitative genetic models for describing simultaneous and recursive relationships between phenotypes. *Genetics* 167:1407–1424.
- Jamrozik, J., J. Bohmanova, and L. R. Schaeffer. 2010. Relationships between milk yield and somatic cell score in Canadian Holsteins from simultaneous and recursive random regression models. *J. Dairy Sci.* 93:1216–1233.
- Jamrozik, J., and L. R. Schaeffer. 2010. Application of multiple-trait finite mixture model to test-day records of milk yield and somatic cell score of Canadian Holsteins. *J. Anim. Breed. Genet.* doi:10.1111/j.1439-0388.2010.00875.x
- Kass, R. E., and A. E. Raftery. 1995. Bayes factors. *J. Am. Stat. Assoc.* 90:773–795.
- Madsen, P., M. M. Shariati, and J. Ødegård. 2008. Genetic analysis of somatic cell score in Danish Holsteins using a liability-normal mixture model. *J. Dairy Sci.* 91:4355–4364.
- Ødegård, J., J. Jensen, P. Madsen, D. Gianola, and B. Heringstad. 2003. Detection of mastitis in dairy cattle by use of mixture models for repeated somatic cell scores: A Bayesian approach via Gibbs sampling. *J. Dairy Sci.* 86:3694–3703.
- Ødegård, J., P. Madsen, D. Gianola, G. Klemendstal, J. Jensen, B. Heringstad, and I. R. Korsgaard. 2005. A Bayesian threshold-normal mixture model for analysis of a continuous mastitis-related trait. *J. Dairy Sci.* 88:2652–2659.
- Schaeffer, L. R., J. Jamrozik, G. J. Kistemaker, and B. J. Van Doormall. 2000. Experience with a test-day model. *J. Dairy Sci.* 83:1135–1144.
- Wu, X.-L., B. Heringstad, Y.-M. Chang, G. de los Campos, and D. Gianola. 2007. Inferring relationships between somatic cell score and milk yield using simultaneous and recursive models. *J. Dairy Sci.* 90:3508–3521.
- Wu, X.-L., B. Heringstad, and D. Gianola. 2010. Bayesian structural equation models for inferring relationships between phenotypes: A review of methodology, identifiability, and applications. *J. Anim. Breed. Genet.* 127:3–15.