



Phenotypic and genetic characterization of novel somatic cell count traits from weekly or monthly observations

J. I. Urioste,*†¹ J. Franzén,* and E. Strandberg*

*Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, PO Box 7023, S-75007 Uppsala, Sweden

†Departamento de Producción Animal y Pasturas, Fac. de Agronomía, UDELAR, Garzón 780, 12900 Montevideo, Uruguay

ABSTRACT

The objectives of this study were (1) to explore traits that better capture weekly or monthly changes in somatic cell counts (SCC) than does the commonly used lactation-average SCC, (2) to estimate their heritabilities and relationships to clinical mastitis (CM), and (3) to determine if these traits are feasible for use in monthly testing schemes. Clinical mastitis and weekly test-day (TD) records of SCC and milk production traits from 1,006 lactations of Swedish Red and Holstein cows collected from 1989 to 2004 were used (data set W). A data subset was also created to mimic monthly recording (data set M, 980 lactations). Twenty SCC traits were defined, taking into account SCC general levels and variation along the lactation curve, time and level of infection, and time of recovery. To reduce dimensionality, cluster and stepwise logistic regression procedures were applied. In data set W, 3 traits, “standard deviation of SCC over the lactation,” a discrete (0/1) indicator of “at least one TD with SCC >500,000 cells/mL,” and “number of days sick in the widest SCC peak” (DWidest) were the variables kept both with cluster procedures and a stepwise logistic regression with the logit of CM as dependent variable. In data set M, DWidest was replaced by “number of SCC peaks” and “average number of days sick per peak” (ADSick). Lactation-average SCC (in the first 150 d or between 150 and 305 d) did not enter into the logistic regression. Heritability estimates obtained for these new traits under a Bayesian setting and a Gibbs sampling approach were 10 to 16% (except for ADSick: 5%). Heritabilities were at least as high in the monthly data set as in the weekly data set. Thus, these SCC traits seem promising for use in breeding programs based on monthly milk recording.

Key words: somatic cell count, genetic parameter, novel trait

INTRODUCTION

Mastitis is an inflammation of the mammary gland associated with considerable adverse economic effects in dairy cattle (Nielsen, 2009). Genetic variation in susceptibility to the disease exists, with heritabilities from 0.07 to 0.12 (e.g., Lund et al., 1999; Heringstad et al., 2003). However, routine recording of clinical mastitis (CM) is not performed in most countries; instead, SCC has been used in genetic evaluations (Mark et al., 2002; Miglior et al., 2005), relying on the positive genetic correlation between SCC and mastitis (e.g., Heringstad et al., 2000; Carlén et al., 2004; Vallimont et al., 2009).

Both mastitis and SCC have complex biological backgrounds (Harmon, 1994). Despite this, the log of lactation-average SCC has often been used as an indicator of the former. A shortcoming of this SCC measure is that the dynamic nature of mastitis is ignored. Clinical mastitis and subclinical (SCM) mastitis cause deviations from a typical curve of SCC. Specific pathogens involved in cases of CM affect the curve differently (de Haas et al., 2004). Use of individual test-day (TD) records of SCC in a random regression model (e.g., Reents et al., 1995) is an advancement compared with lactation average, but the polynomial model is not well suited for the drastic changes in SCC associated with cases of (especially) CM. Therefore, it could be worthwhile to explore alternative traits derived from the SCC curve to improve possibilities of genetic progress in selection against CM. The new traits should be flexible enough to accommodate for sudden and drastic changes in SCC, especially resulting from a case of CM.

Alternative approaches have been attempted. A baseline curve corresponding to the level of the healthy cow (in a certain group, e.g., parity) has been modeled, with an optional curve describing changes corresponding to incidence of mastitis; that is, as a deviation from the baseline (de Haas et al., 2002). Patterns of increase and decrease in the SCC curve might be related to the type of mastitis, clinical or subclinical, and to disease resistance once an infection has occurred (de Haas et al., 2004). Distribution characteristics (standard deviation of the TD SCC, maximum and minimum values) have

Received March 29, 2010.

Accepted September 1, 2010.

¹Corresponding author: jurioste@fagro.edu.uy

also been considered (Green et al., 2004), although not for genetic studies. Another option has been to develop mixture models to better account for some hidden record structure due to unknown disease status (Deteillex and Leroy, 2000; Ødegård et al., 2003; Boettcher et al., 2007; Madsen et al., 2008). The latter methods are theoretically attractive but currently difficult to use in a practical setting for prediction of breeding values because of computational limitations. Therefore, the focus of this study was on novel traits that could be derived from information present in the TD SCC records, to be used as indicator traits for mastitis.

The objectives of this study were to (1) explore a variety of novel traits that capture changes along the SCC curve; (2) characterize phenotypic relationships among them, and describe their relationship with CM; (3) identify genetic variation in these traits; and (4) extend the results obtained from a research herd data set with records once a week to a subset of the same data with monthly intervals, as is common in field data.

MATERIALS AND METHODS

Swedish Animal Care and Use Committee approval was not needed for this study because the data were extracted from an existing database.

Data

Data were TD records collected at weekly intervals in the Jälla research herd of the Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, near Uppsala, Sweden. Full details can be found in Hagnestam et al. (2007). Data sampled between November 1989 and March 2004 were available. Cows were of the Swedish Red (**SR**, 648 lactations) and Swedish Holstein (**SH**, 358 lactations) breeds, were milked twice daily, and had an average yearly production of 8,900 and 10,600 kg of milk, respectively. Swedish Red cows were further split into high (**SR-H**) or low (**SR-L**) fat content selection lines. Cases of CM were detected by the milkers, based on the presence of abnormal milk in the first milk streams or by signs of inflammation in one or more udder quarters. All cases were diagnosed by a veterinarian.

Two data sets were created to investigate the effects of using weekly or monthly records in SCC traits. Data set **W** (weekly measures) included 1,006 cow-lactation records (434 first parity and 572 second or later parities, the average parity number per cow being 2.06) with a minimum of 10 TD present both in the first (≤ 150 DIM) and second half of the lactation (> 150 DIM), obtained from 467 cows (290 had repeated records, of which 269 were primiparous), sired by 147 bulls (36,051 TD ob-

servations in total). Data set **M** (monthly measures) mimicked a field data set by keeping one observation from the complete data set by approximately every 30 d. The data set contained 980 cow-lactation records (with a minimum of 5 TD spread over early and late lactation periods) originated from 9,159 TD observations from 457 cows sired by 147 bulls.

Traits

Clinical mastitis was scored as present (1) in a given lactation if at least one case of CM was recorded; otherwise, it was scored as absent (0). Various alternative SCC traits were defined as potential indicator traits for CM. The new traits were designed to capture SCC general levels and variation along the curve, time, and level of infection, and time of recovery. Production traits (e.g., milk yield and fat to protein ratio) were included as a reference.

A detailed definition of the used traits follows:

- For simplicity, it was assumed that CM, as defined above, was the same trait along the lactation (DIM 1 to 305), because most first cases of CM in the research herd occurred in the first lactation week (Hagnestam et al., 2007); this pattern has often been reported in the literature (Emanuelson et al., 1988; Barkema et al., 1998; de Haas et al., 2002).
- Test-day SCC were averaged over different lactation periods (5–150 d, 151–305 d) to compare the first versus second parts of the lactation (**SCC150D** and **SCC150–305D**, respectively), as in the approach by de Haas et al. (2008).
- As suggested by Heuven (1987), TD-SCC was multiplied by TD milk yield to obtain total number of cells and then log-transformed (**LTSCC**), to circumvent possible effects of dilution.
- It was further hypothesized that the dynamics of SCC during lactation differs between healthy and diseased mastitis cows. Following de Haas et al. (2002, 2004, 2008), we assumed that a test-day SCC exceeding a given threshold could originate from a mastitis infection, but did not define the threshold. Instead, a grid of different threshold values (40, 80, 150 and 500 thousand cells/mL) was applied for TD-SCC, trying to identify TD with increasing probability of infection (or inflammation), and particularly those with higher severity. Accordingly, each TD-SCC was classified individually as belonging to a certain SCC interval. Binary traits were defined consequently, depending on TD classification, as follows: the presence of at least one TD with <

40,000 cell/mL (TD<40), at least one TD between 41,000 and 80,000 cells/mL (TD41–80), at least one TD between 81,000 and 150,000 cells/mL (TD81–150), at least one TD between 151,000 and 500,000 cells/mL (TD151–500) or at least one TD >500,000 cells/mL (TD>500). As an example, a TD-SCC of 230,000 cells/mL should result in a recorded “1” for the trait TD 151–500 for that animal and lactation.

- e) Furthermore, number of test-days in the lactation with SCC in each threshold interval were also considered as traits (NTD<40, NTD41–80, NTD81–150, NTD151–500, and NTD>500). No correction for shorter lactations was made; instead, a restriction was imposed on the data, such as to guarantee the presence of test-days both in early and late lactation.
- f) An infection peak was defined as a period of increased SCC (>150,000 cells/mL) between 2 low (\leq 150,000 cells/mL) TD observations. The number of peaks (**NPeak**) was considered as a trait.
- g) Patterns of SCC peaks could also be distinguished based on number of days of increased SCC during lactation. Thus, a practical approximation of length of disease was recorded as the number of days between the start and the end of an infection peak (TD with cell counts <150,000 cells/mL and having at least a TD>150,000 in between). Three alternative traits were defined: number of days in the widest peak (**DWidest**), counting only days in the widest peak; total number of days sick (**DSick**), summing up all days in each peak; and average days sick per peak (**ADSick**), as DSick divided by NPeak. In line with de Haas et al. (2008), traits were not corrected for a baseline effect, as suggested earlier by the same authors (de Haas et al., 2002), in an attempt at simplification.
- h) Green et al. (2004) showed that high maximum value and standard deviation log SCC during lactation (**SCCSD**) were the best overall indicators of CM. The latter measure was added in the present study.

A list of the 20 SCC traits considered in this study is presented in Table 1. Two production traits, average test-day milk (**AVMilk**) and average fat:protein ratio (**AVFPR**), were also used as references.

Traits that showed closest association with CM in the phenotypic analyses were subsequently examined for their genetic background, together with the reference production traits. Some of them (CM, TD41–80, TD151–500, TD>500) were treated as threshold traits,

whereas AVMilk, AVFPR, SCCSD, DWidest, and ADSick were considered as continuous; NPeak was treated both as a threshold (0; 1–2; \geq 3) and as a continuous trait. The traits SCCSD, DWidest, ADSick, and NPeak were transformed using a natural logarithm to improve normality; DWidest and ADSick, which depend on the presence of a peak, were treated as missing values in the case of a nonexisting infection peak.

Methods

Clustering Novel SCC Traits. Considering that all traits are based on the same original information (TD-SCC), it is likely that some of the created traits contain more or less the same information, and that they are highly correlated. Because of the large number of developed traits, we wanted to narrow the field of potentially interesting traits. One approach is to find the variables that best gather the observations into a certain number of clusters. The newly created SCC traits (Table 1) were standardized to mean zero and unity SD. A first approach was done to decide the number of clusters (Proc Cluster of SAS, with Ward's minimum variance method and pseudo F statistics; SAS Institute Inc., Cary, NC). Clustering of SCC variables was performed with the k-means approach (Proc Fastclus of SAS). Canonical discriminant analysis was then applied to identify and select a subset of variables that explains most of the differences between clusters (Proc Candisc and Proc Stepdisc of SAS). Final clusters were defined, based on chosen main variables (Proc Fastclus of SAS).

Logistic Regression. Somatic cell count is used mainly to predict occurrence of mastitis. We were interested in testing how well the different variables coming out from the cluster analysis could predict CM. A logistic linear regression model was implemented through Proc Logistic of SAS, where the logit of CM was the response variable and some of all possible candidate SCC traits were selected as explanatory variables in a stepwise procedure. Effects of parity (primiparous and multiparous cows) and breed (SH, SR-H and SR-L) were also included in the procedure and found to be significant. Previous runs with data set M showed illogical regression coefficients in some variables. With a closer look, we detected collinearity between the variables TD>500 and TD41–80, and between TD>500 and TD151–500, because of negative (-0.29) and positive (0.30) correlations, respectively ($P < 0.05$). As a result, either variable captured largely the effect of the others. In successive exploratory runs, these variables were not present simultaneously in the same model; the final model arbitrarily excluded the variable TD151–500 from the stepwise procedure.

Table 1. Variable acronyms, descriptions, and average, minimum, and maximum values for the weekly (W) and monthly (M) data sets

Variable acronym	Description	Data set W			Data set M		
		Average	Minimum	Maximum	Average	Minimum	Maximum
SCC general levels/ time of infection							
SCC150D	Average test-day (TD) SCC between 5 and 150 DIM, d	160,776	8,000	1,927,240	129,633	9,600	2,324,200
SCC150–305D	Average TD SCC between 151 and 305 DIM, d	207,597	11,389	11,389	178,351	12,000	2,198,000
LTSCC	Log of total amount of SCC in milk	21.41	19.55	24.39	21.30	19.51	24.18
DIMSCC > 150	Average DIM until SCC observations >150,000 cell/mL, d	134.72	5	305	124.35	5	303
Variation in SCC curve							
SCCSD	SCC standard deviation, cell/mL	220,671	4,398	2,349,590	154,998	4,193	2,623,190
Level of infection/ inflammation							
TD<40	At least one TD SCC ≤40,000 cell/mL (0,1)	0.77	0	1	0.73	0	1
TD41–80	At least one TD SCC between 41,000 and 80,000 cell/mL (0,1)	0.90	0	1	0.76	0	1
TD81–150	At least one TD SCC between 81,000 and 150,000 cell/mL (0,1)	0.89	0	1	0.65	0	1
TD151–500	At least one TD SCC between 151,000 and 500,000 cell/mL (0,1)	0.81	0	1	0.59	0	1
TD>500	At least one TD SCC >500,000 cell/mL (0,1)	0.50	0	1	0.24	0	1
NTD<40	No. of TD with SCC ≤40,000 cell/mL	12.35	0	39	3.89	0	10
NTD41–80	No. of TD with between 41,000 and 80,000 cell/mL	7.24	0	30	1.99	0	7
NTD81–150	No. of TD with between 81,000 and 150,000 cell/mL	5.92	0	31	1.63	0	8
NTD151–500	No. of TD with between 151,000 and 500,000 cell/mL	6.91	0	36	1.95	0	10
NTD>500	No. of TD with SCC >500,000 cell/mL	2.29	0	34	0.54	0	10
NTD>150	No. of TD with SCC >150,000 cell/mL	9.20	0	41	2.50	0	10
NPeak	No. of peaks (SCC >150,000 cell/mL)	2.46	0	10	0.92	0	4
Time of recovery							
DWidest	No. of days sick in the widest peak	56.34	0	300	71.16	0	300
DSick	Total no. of days sick (sum of number of days sick over peaks)	92.00	0	300	87.74	0	300
ADSick	Average days sick per peak	36.42	0	300	64.29	0	300
Production traits							
AVMilk	Average TD milk yield, kg	27.68	14.85	50.85	27.31	14.92	47.82
AVFPR	Average fat:protein ratio	1.26	0.79	1.65	1.26	0.89	1.62
Mastitis							
CM	Presence (1) or absence (0) of clinical mastitis (CM) during the lactation	0.230	0	1	0.215	0	1

Genetic Models. Single-trait analyses were run for the candidate traits identified in earlier steps, using the following general mixed animal model:

$$y_{ijklm} = \mu + byr_i + br_j + lact_k + a_l + pe_l + e_{ijklm}, [1]$$

where y_{ijklm} denotes the response trait; μ is a general mean; byr_i is birth-year group i (3 levels: 1985 to 1990, 1991 to 1995, and greater than 1995); br_j is breed (3 levels: SH, SR-H, and SR-L); $lact_k$ is lactation number (5 levels, first to fifth or more); a_l is the effect of breeding value of animal l with assumption $a_l \sim N(0, \mathbf{A}\sigma_A^2)$,

where \mathbf{A} is the additive relationship matrix and σ_A^2 the additive genetic variance; pe_l is permanent effect of repeated cow measurements (467 cows in data set W and 457 in data set M), with $pe_l \sim N(0, \sigma_{pe}^2)$; and e_{ijklm} is the residual, $e_{ijklm} \sim N(0, \sigma_e^2)$.

Animal effects (1,717 animals from the pedigree file) were linked using the relationship matrix. A threshold liability approach (e.g., Gianola and Foulley, 1983) was used for traits expressed as a discrete response. In the binary case, the threshold model postulates an underlying continuous random variable, called liability, λ , such that the observed responses are $y_i = 0$ if $\lambda_i \leq \tau$, or $y_i = 1$ if $\lambda_i > \tau$; where τ is a fixed threshold and $y_i = 1$ or 0 corresponds to the observed presence or absence of the trait for observation i , respectively. For these traits, λ replaced y in equation [1]. Because the threshold and σ_e^2 are not identifiable, these parameters are usually set to some arbitrary values ($\tau = 0$ to denote origin and $\sigma_e^2 = 1$ to define the scale of measurement). Extension to more than 2 outcomes is straightforward.

Additive genetic, permanent, and residual variances were drawn from the posterior distributions using a Bayesian approach and Gibbs sampling, as implemented for threshold or linear trait analyses in the program Thrgibbs1f90 (Misztal et al., 2002). Based on visual inspection of trace plots in earlier runs (a binary and a continuous variable were tested with chains of 100,000, 250,000, and 500,000 iterations, thinning intervals of 50 and 100 samples, and 20% of burn-in in all cases, and found to converge to the same values), a chain of 120,000 iterations was run for each trait, with a burn-in of 20,000 rounds, keeping every 50th sample for inference of posterior features (2,000 samples). Heritability (h^2) was computed in the usual way, as the ratio between the additive genetic and the phenotypic variances. The repeatability (r) was calculated as the ratio between the additive genetic plus permanent environment variance components and the phenotypic variance. Post-Gibbs

analysis (posterior means, medians, percentiles, distributions) was done with SAS (SAS Institute Inc.).

RESULTS

Overall mastitis rate in data sets W and M were 23 and 21.5%, respectively; 81 and 61.2% of lactations had at least one TD with SCC >150,000 cell/mL, and 49.9 and 25.7% had at least one TD with SCC >500,000 cell/mL in data sets W and M, respectively. Infection peaks in data set W were detected in 849 lactations (84.4% of total), with 754 lactations having between 1 and 5 peaks, and 95 having 6 or more. For cows with infection peaks, mean total number of days sick was 109 d, with an average of 43.2 d per peak. In data set M, 636 lactations with peaks, of 980 in total, were identified (64.9%). Because the measurements are taken once a month, a good part of the original information contained in data set W is lost. Hence, the maximum number of peaks was only 4, with most observations (413) showing only 1 peak during lactation. The mean total number of days sick was 134.3 d and average value per peak was 100.4 d.

Phenotypic Associations

Clustering. Evaluation of preliminary runs indicated that a reasonable number of clusters should lie between 2 and 4. After inspection of results, we decided to concentrate on 3 clusters; other results are not reported. To simplify interpretation of the analyses, a discriminant procedure identified a subset of variables that explained most of the differences between clusters; variables are shown in Table 2 (data set W) and Table 3 (data set M).

A measure of SCC curve variation, the SCCSD, was the variable entered first to explain variation between clusters in both data sets. Traits indicating general levels of SCC (Table 1) were not selected; instead, discrete (0, 1) measures of SCC depicting level of infections (TD>500 in both sets and TD41–80 and TD151–500 in data set M) were chosen. Additionally, variables illustrating the dynamics of infection, such as the patterns of infection peaks (DWidest in data set W and NPeak in data set M) were selected. The production variable associated most closely with differences between clusters was fat:protein ratio.

Results from the cluster analysis using data set W are given in Table 2 and Figure 1. Cluster 1 was associated with lower SCCSD observations (52,100 cells/mL), absence of severe infection peaks, shorter periods of sickness (a maximum of 24.5 d, suggesting quick recovery), and moderate fat:protein ratio. Cluster 3 had opposite features, showing high variation, presence of severe

Table 2. Means for the variables clustered for weekly data (data set W) and number of mastitic and nonmastitic cows in each cluster

Variable ¹	Cluster 1 (n = 503)	Cluster 2 (n = 365)	Cluster 3 (n = 138)
SCCSD	52.1	289.4	652.7
TD>500	0	1.0	1.0
DWidest	24.5	52.6	182.3
AVFPR	1.27	1.27	1.23
Number (%) of cows with clinical mastitis (n = 231)	33 (7) ²	135 (37) ²	63 (46) ²
Number (%) nonmastitis cows (n = 775)	470 (93)	230 (63)	75 (54)

¹SCCSD = SCC standard deviation, cell/mL; TD>500 = at least one test-day (TD) SCC >500,000 cell/mL (0,1); DWidest = no. of days sick in the widest peak; AVFPR = average fat:protein ratio.

²Percentage of respective cluster.

peaks, wide infection peaks, and lower fat:protein ratio. Cluster 2 was in between: severe peaks were present, but SCC variation and width of infection peaks were intermediate. A graphic illustration of clusters using 2 canonical variables is in Figure 1.

Based on these results, one could hypothesize that most healthy cows should be found in cluster 1, whereas mastitic cows should be more often associated with cluster 3; cluster 2 could gather cows in an intermediate situation (e.g., recovering from the disease). In fact, most healthy, nonmastitic cows (470 in 775) were associated with cluster 1 in data set W (Table 2). In total, 93% of the cows (470 in 503) assigned to cluster 1 were nonmastitic. The relationship of mastitic cows with the clusters was, however, less clear: only 46% of the cows (63 in 138) in cluster 3 were mastitic, and 135 out of 231 mastitis cows (58%) were grouped within cluster 2. In general, a non-negligible number of false-positive and false-negative observations were found. Misclassification was partly associated with the timing of mastitis diagnosis: median value of diagnosis date, calculated in days from calving was 3 d for mastitis cows belonging

to cluster 1 (“healthy” cows), 71 d for cluster 2 and 91 d for cluster 3 (“diseased” cows).

Similar trends were observed when using data set M (Table 3 and Figure 2), although here different variables (TD41–80, TD151–500, NPeak) were selected by the procedure instead of DWidest. Fewer observations were assigned to cluster 3, however, which can be interpreted as a difficulty to identify such animals because of less information contained in this data set. As a result, a very large variation in SCC, low presence of SCC TD between 41,000 and 80,000 cells/mL (“normal” SCC levels), presence of both moderate and severe test-day SCC records, and low fat:protein ratio was observed. Likewise, within-cluster variation was larger compared with that of data set W, as illustrated by Figure 2. Furthermore, 65% of cows in cluster 3 were mastitic (26 cow-lactations with CM in a cluster with 40 observations), but 88.6% of mastitic cows were distributed within clusters 1 and 2, which expresses an important loss of efficiency when attempting to relate the SCC traits with CM compared with data set W. Following the same pattern as in data set W, the median value of

Table 3. Means for the variables clustered for monthly data (data set M) and number of mastitic and nonmastitic cows in each cluster

Variable ¹	Cluster 1 (n = 565)	Cluster 2 (n = 375)	Cluster 3 (n = 40)
SCCSD	50.8	238.2	1,355.1
TD41–80	0.88	0.53	0.27
TD151–500	0.39	0.92	0.80
TD>500	0	0.56	1.0
NPeak	0.39	1.60	1.45
AVFPR	1.27	1.25	1.23
Number (%) of cows with clinical mastitis (n = 221)	79 (14)	116 (31)	26 (65)
Number (%) of nonmastitis cows (n = 759)	486 (86)	259 (69)	14(35)

¹SCCSD = SCC standard deviation, cell/mL; TD41–80 = at least one TD SCC between 41,000 and 80,000 cell/mL (0,1); TD151–500 = at least one TD SCC between 151,000 and 500,000 cell/mL (0,1); TD>500 = at least one TD SCC >500,000 cell/mL (0,1); NPeak = No. of peaks (SCC >150,000 cell/mL); AVFPR = average fat:protein ratio.

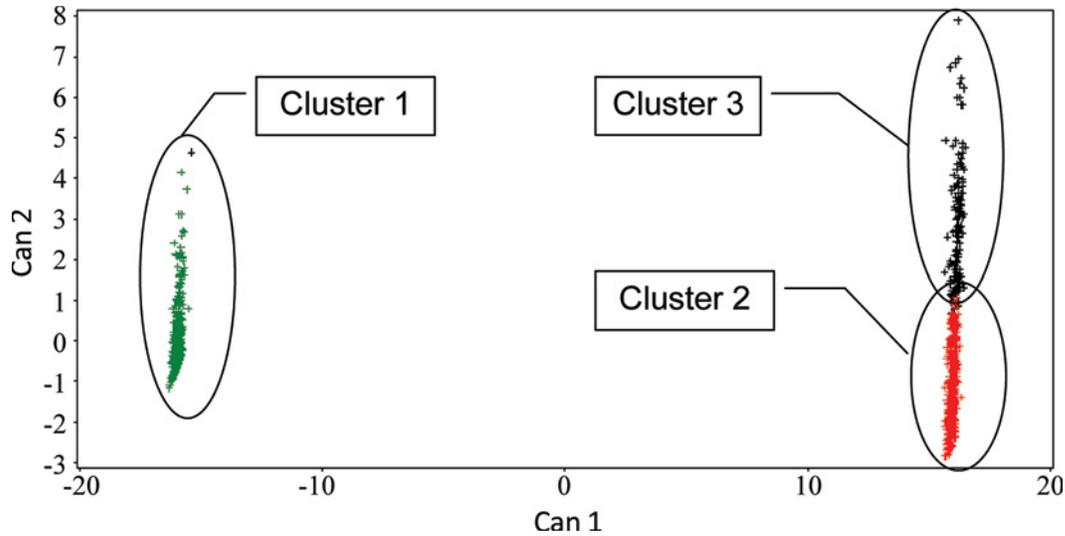


Figure 1. Illustration of clusters in data set W (weekly data), using the canonical (Can) variables 1 and 2 on the x- and y-axes, respectively. Color version available in the online PDF.

diagnosis date was 5 d after calving for mastitis cows in cluster 1, 81 d for cluster 2, and 137 d for cluster 3.

Logistic Regression. Results from the logistic regression are shown in Table 4. The effects of breed, lactation, TD>500, and SCCSD were significant in both data sets. Incidence of CM was lower in primiparous than in multiparous cows in both data sets. Furthermore, the SH breed was associated with higher incidence of CM; the lowest incidence was for SR-H, with intermediate values for SR-L. In data set W, the variable DWidest entered the model, whereas NPeak and ADSick did so for data set M. As in the cluster

analysis, standard measures of average SCC (i.e., SCC150D and SCC150–305D) were not among the variables that explained most of the variation in CM. The variables chosen using either cluster analysis or logistic regression were partially the same, and therefore we concluded that they should be checked first for genetic variability.

Genetic Variation

In spite of the limitations posed by the data set available, evidence of additive genetic variability was found

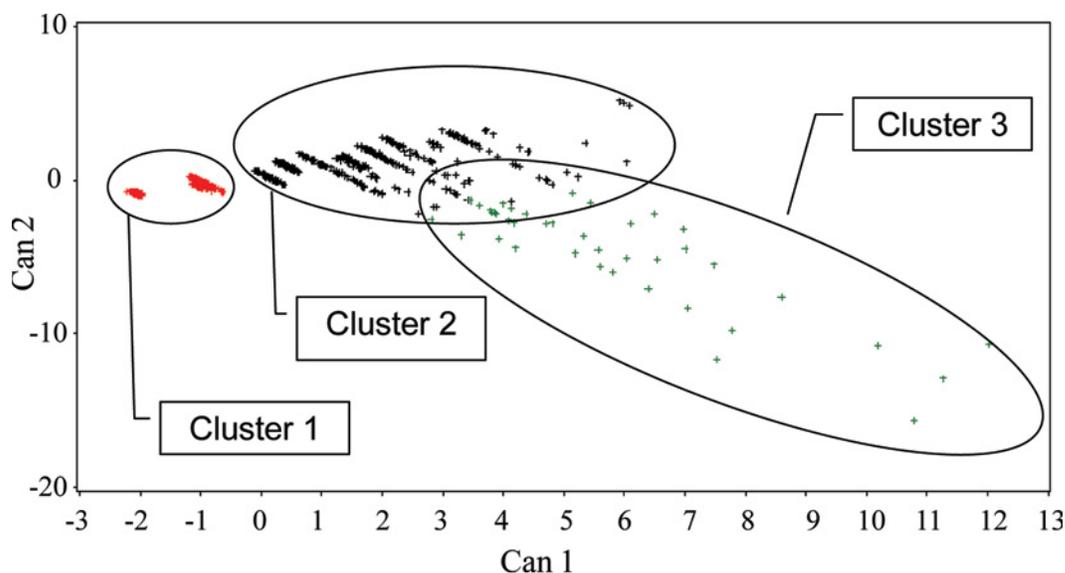


Figure 2. Illustration of clusters using data set M (monthly data), using the canonical (Can) variable 1 on the x-axis and canonical variable 2 on the y-axis. Color version available in the online PDF.

Table 4. Maximum likelihood (ML) estimates in a logistic stepwise regression for clinical mastitis for weekly data (data set W) and monthly data (data set M)¹

Variable ²	W data set		M data set	
	ML estimate	Pr > chi ³	ML estimate	Pr > chi ³
Intercept	-2.7013	<0.0001	-1.9614	<0.0001
Breed				
Swedish Holstein	0.4036	0.0009	0.3643	0.0040
Swedish Red-High fat line	-0.3304	0.0136	-0.4553	0.0009
Lactation				
1	-0.3026	0.0017	-0.1574	0.0788
SCCSD	0.0031	<0.0001	0.0020	<0.0001
TD>500	1.5647	<0.0001	0.5900	0.0214
DWidest	-0.0072	<0.0001	—	—
NPeak	—	—	0.3401	<0.0016
ADSick	—	—	-0.0040	<0.0065

¹Variables in the table are those that were entered into the model.

²SCCSD = SCC standard deviation, cell/mL; TD>500 = at least one TD SCC >500,000 cell/mL (0,1); DWidest = no. of days sick in the widest peak; NPeak = No. of peaks (SCC >150,000 cell/mL); ADSick = average days sick per peak.

³Significance of the variable entering the model, using a chi-squared test statistic.

for all novel SCC traits (Table 5). Posterior heritability values ranged from 0.05 to 0.16 estimated from data set W and 0.05 to 0.22 for those estimated from data set M. The lowest estimate of heritability was that of the average days sick (0.05 in both data sets). Most estimates were larger than the heritability for CM; that is, 0.04 in both data sets. A general measure of SCC such as average test-day SCC in early lactation (SCC150D) was added for comparison purposes, and its heritability was similar to that of other SCC traits (0.09 to 0.12). As expected, production traits showed higher heritability estimates: 0.23 for milk yield and 0.52 to 0.58 for fat:protein ratio. In general, no large differences were observed between data sets W and M in parameter estimates. Uncertainty in all estimates, defined as the 90% credibility interval between P5 and P95 percentiles, was large.

Repeatabilities were also obtained (Table 5). Again, estimated values for the alternative SCC traits were higher (0.12 to 0.36 in data set W and 0.17 to 0.33 in data set M) than those for CM (0.09 to 0.11). Repeatabilities for production traits were very high: 0.50 to 0.51 for milk yield and 0.71 to 0.77 for fat:protein ratio.

DISCUSSION

Somatic cell count is a useful measure of udder health, because it is widely recorded and genetically associated with CM. One of its potential shortcomings as indicator trait is that acute, short-duration infections may be difficult to identify simply from increased average SCC during lactation, because SCC is often recorded at approximately monthly intervals. It is ap-

parent that different mastitis pathogens elicit different SCC responses; in particular, CM associated with environmental organisms such as *Escherichia coli* may result in a high SCC for a short period (de Haas et al., 2002; 2004).

The main objectives of this study were to acquire knowledge about SCC traits other than lactation-average SCC using data with weekly observations, by characterizing phenotypic relationships among them and with CM, identifying genetic variation in these traits, and assessing the consequences of utilizing monthly instead of weekly SCC records. The experimental herd data analyzed in this paper have limitations in structure, size, and the effect of herd-specific management, which reduces the possibility of making precise inferences about genetic parameters. The production levels in the herd were slightly above national averages (see Hagnestam et al., 2007 for detailed description). Moreover, compared with Swedish average figures, the herd has low bulk tank SCC but high incidence of CM, most likely because the surveillance of the animals was probably more extensive than what would be expected in commercial dairy herds and more cases of CM are likely to be detected (Nielsen, 2009). These disadvantages are counteracted by the more consistent definition of CM and the more frequent SCC testing. Given the paucity of information regarding new SCC traits in dairy cattle, the analysis conducted provided some good insight into the attributes of various SCC traits as suitable variables for selection.

An important finding in this study is that traits such as SCCSD and presence of TD with SCC >500,000 cells/mL were selected in first place when trying to explain phenotypic relationships with CM. Both traits

Table 5. Posterior median (percentiles 5 and 95 within parentheses) values of heritabilities (h^2) and repeatabilities (r) from data sets with weekly (W) and monthly (M) records of clinical mastitis (CM) and SCC

Variable ¹	h^2		r	
	Data set W	Data set M	Data set W	Data set M
CM	0.04 (0.01–0.11)	0.04 (0.00–0.11)	0.09 (0.03–0.17)	0.11 (0.04–0.20)
TD41–80	0.15 (0.04–0.32)	0.08 (0.01–0.17)	0.36 (0.21–0.51)	0.24 (0.14–0.35)
TD151–500	0.16 (0.07–0.30)	0.22 (0.11–0.33)	0.36 (0.25–0.46)	0.33 (0.26–0.44)
TD>500	0.12 (0.03–0.24)	0.12 (0.04–0.24)	0.34 (0.25–0.43)	0.32 (0.22–0.42)
SCCSD ²	0.10 (0.03–0.19)	0.13 (0.06–0.21)	0.25 (0.19–0.32)	0.26 (0.19–0.33)
NPeak	0.12 (0.04–0.24)	0.16 (0.07–0.26)	0.28 (0.20–0.37)	0.29 (0.20–0.38)
NPeak ²	0.10 (0.04–0.16)	0.12 (0.04–0.20)	0.18 (0.12–0.25)	0.24 (0.17–0.31)
DWidest ²	0.13 (0.04–0.22)	0.14 (0.08–0.21)	0.31 (0.24–0.37)	0.25 (0.18–0.32)
ADSick ²	0.05 (0.01–0.11)	0.05 (0.01–0.13)	0.12 (0.07–0.18)	0.24 (0.17–0.30)
AVMilk	0.23 (0.09–0.37)	0.23 (0.11–0.36)	0.51 (0.45–0.57)	0.50 (0.44–0.56)
AVFPR	0.58 (0.45–0.70)	0.52 (0.38–0.64)	0.77 (0.74–0.80)	0.71 (0.66–0.75)
SCC150D	0.09 (0.02–0.17)	0.12 (0.06–0.19)	0.23 (0.17–0.29)	0.24 (0.17–0.30)

¹See Table 1 for trait definitions.

²Variable was transformed using a natural logarithm before analysis.

showed sizeable genetic variation, making them good candidates for genetic selection against mastitis.

Phenotypic Associations

A distinguishable relationship was observed among SCC variables and between them and CM. Applying multivariate techniques, we have shown that a cluster with mainly healthy animals tended to have relatively low SCC variation, without severe infection peaks; when sick, these animals had a quicker recovery (less than a month). Most nonmastitic cows belonged to this cluster of animals. At the opposite extreme, a cluster of cows with remarkably high SCC variation along the lactation, as well as features of infection peak patterns, such as long-lasting sick periods (approximately 6 mo) or higher number of infection peaks, was observed. Almost half of the cows in this cluster had CM. The length of disease calculated in our study is more likely a measure of number of days with high TD-SCC, and probably more related to events of subclinical mastitis; it certainly overestimates the real number of CM sick days because of the definition employed, and must be interpreted with caution. Here, it is only used as a relative value. The presence of patterns of peaks in lacta-

tions has been shown to be phenotypically informative for pathogen specific CM (de Haas et al., 2004). The strong association between SCCSD and mastitis is also in accordance with the phenotypic relationships found by Green et al. (2004).

One of the limitations of SCC in general is that the cow must be in milk before the measure is available. Commonly, the cow must be 5 d postpartum before it is milked in the recording scheme. With monthly tests, the average first test-day would be at roughly 20 d postpartum. It was observed that mastitic cows classified within the cluster of healthy cows had a very early mastitis diagnosis (on average, 3 d after calving in data set W and 5 d in data set M), which could explain, in part, some of the misclassifications (mastitic cows classified within a cluster where most cows were healthy) reported here.

Multivariate techniques to reduce variable dimensions are scarce in animal breeding applications. Ibáñez et al. (1996) applied a hierarchical cluster analysis in dairy data to define classes containing contemporary groups with similar variability, thus reducing 36,781 contemporary groups into 38 phenotypic standard deviation classes. Weigel and Rekaya (2000) clustered herds from 13 countries, regardless of borders, using variables re-

lated to management, climate, and genetic composition of the herds. Each variable was standardized, transformed into principal components, and then retained if eigenvalues were ≥ 0.50 . Further application of this methodology is found in Zwald et al. (2003). Leclerc et al. (2005) used a principal components approach to reduce the number of parameters to be estimated in international genetic evaluations for dairy cows, and Soy-eurt et al. (2010) estimated genetic parameters of the mid-infrared milk spectrum, reducing 1,060 data points per sample into 46 principal components, suggesting the possibility of finally considering only 8 traits. This approach can be a powerful tool for characterizing, for instance, phenotypic variation, and it is of particular interest for studying scenarios such as ours, where a small subset of the variables can explain a large part of this variation.

Logistic regression is a common statistical approach to deal with phenotypic analysis of mastitis in dairy cows (e.g., Waage et al., 1998; Green et al., 2002; Pantoja et al., 2009), taking into account the binary nature of the trait. In our case, the results obtained reinforced the findings of the cluster analysis, showing a strong relationship between CM and SCCSD and TD >500 ; more sick periods, represented by the number of infection peaks, were also linked to CM. Surprisingly, both DWidest and ADSick were inversely, although weakly, related to CM (Table 4), implying that a longer sick period, either for the widest SCC peak or on average, decreased the risk of CM. This could be simply due to a collinearity effect; another possible biological interpretation could be that clinical cases (all those reported in the present study were of this kind) were shorter in duration partly because they were actually treated. Another possible interpretation is that cows with longer sick periods are subclinically infected but never become clinical cases.

The fat:protein ratio was a factor included when clustering the observations but that did not enter into the logistic regression. A decrease in fat and no major changes in total protein is expected with elevation of SCC (Harmon, 1994). In the study by Windig et al. (2005), fat:protein ratios were nonlinearly related to SCC peak rates: both a lower and a higher fat:protein ratio at the start of lactation increased the risks of SCC peaks. Windig et al. (2005) also found that both an increased and a decreased ratio were associated with an increased risk of mastitis. A high fat percentage at the start of lactation and a lower fat percentage later in lactation are related to a negative energy balance (de Vries and Veerkamp, 2000; Stoop et al., 2009). Thus, a negative energy balance might play a role in the relationship between fat:protein ratio and mastitis.

Genetic Variation

Studies on genetic variability of new SCC traits are very scarce in the literature. One of the difficulties found by the researchers is that SCC is usually recorded at monthly intervals, and responses may be apparent during shorter periods. The work by de Haas et al. (2003; 2008) has shown that traits similar to those developed here are heritable and show sizeable genetic correlations with CM, which in many cases were stronger than that between CM and average SCS. Others (e.g., Green et al., 2004) have defined new traits but did not include genetic analyses. This prior knowledge was applied in a unique data set and confirmed that traits mentioned in the literature plus those added in the present study show genetic variation, independently of recording frequency. The heritabilities estimated in the current investigation were in the range of 0.05 to 0.22, which is consistent with the estimates previously mentioned obtained from monthly TD records. The main finding was that estimates were not consistently affected by using either the W or the M data set, indicating that the suggested traits could be used on a data set with monthly records without great loss of information for selection purposes. The data sets used in this study showed a consistent picture: higher heritabilities were found for the new SCC traits compared with estimates for CM (Table 5), indicating good possibilities for selection. The practical implications are important, because it suggests the feasibility of selecting for reduced mastitis effects in dairy cows.

Because of the small data set, the uncertainty in the estimates was large and genetic correlations could not be properly estimated in this study. Furthermore, some biases generated by the data structure (e.g., cows culled due to udder health problems) could be expected. Therefore, the next step should be to study these traits further using national databases for more accurate estimates of genetic parameters and to verify that similar heritabilities can be found. It is hypothesized that a combination of several definitions of SCC (indicator traits) might be necessary to capture the full scope of the mastitis complex. Then, genetic selection on, say, an index with SCC traits could be proposed, and greater progress might be obtained if the index is included in the breeding program. Progress in this direction has already been reported. de Haas et al. (2008) suggested that some additional information could be gained from the patterns of peaks for CM. In contrast, they found that genetic correlations between subclinical mastitis and patterns of peaks in SCC were weaker than those with lactation-average SCS, indicating that genetic selection for decreased presence of SCC peaks

would not affect the occurrence of SCM as effectively as selecting for lower lactation-average SCS; however, there is always a problem that SCM is defined based on SCC, which should lead to a certain amount of autocorrelation. Eding et al. (2009) implemented estimation of breeding values for 5 indicator traits derived from TD SCC, and Windig et al. (2010) examined combinations of alternative traits in an index that best can reduce mastitis traits. Future studies in this direction are warranted.

CONCLUSIONS

In this paper, some SCC-based traits were considered that have not been commonly used for genetic purposes previously. They were found to be significant phenotypic predictors of CM, to have a reasonable heritability (5–22%), and to be identifiable in both research and (mimicked) field data. Particularly original with respect to earlier studies is the inclusion of measures of number of days sick (the “base” of the SCC peaks) and variation of SCC, and the clustering and stepwise logistic regression approaches, which helped to reduce the high dimensionality of candidate traits. The implication of this research is that it suggests some alternative traits that can be derived from already existing information in routine test-day recordings of SCC. Making better use of that information may therefore improve genetic programs aimed to select for lower susceptibility to CM.

ACKNOWLEDGMENTS

This work was carried out as part of the RobustMilk Project, which 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. We are grateful to the present and past staff of our research herd and the Jälla agricultural school (near Uppsala, Sweden) for the collection of the data.

REFERENCES

- Barkema, H. W., Y. H. Schukken, T. J. Lam, M. L. Beiboer, H. Wilmsink, G. Benedictus, and A. Brand. 1998. Incidence of clinical mastitis in dairy herds grouped in three categories by bulk milk somatic cell counts. *J. Dairy Sci.* 81:411–419.
- Boettcher, P. J., D. Caraviello, and D. Gianola. 2007. Genetic analysis of somatic cell scores in US Holsteins with a Bayesian mixture model. *J. Dairy Sci.* 90:435–444.
- Carlén, E., E. Strandberg, and A. Roth. 2004. Genetic parameters for clinical mastitis, somatic cell score, and production in the first three lactations of Swedish Holstein cows. *J. Dairy Sci.* 87:3062–3070.
- de Haas, Y., H. W. Barkema, Y. H. Schukken, and R. F. Veerkamp. 2003. Genetic associations for pathogen-specific clinical mastitis and patterns of peaks in somatic cell count. *Anim. Sci.* 77:187–195.
- de Haas, Y., H. W. Barkema, and R. F. Veerkamp. 2002. The effect of pathogen-specific clinical mastitis on the lactation curve for somatic cell count. *J. Dairy Sci.* 85:1314–1323.
- de Haas, Y., W. Ouweltjes, J. ten Napel, J. J. Windig, and G. de Jong. 2008. Alternative somatic cell count traits as mastitis indicators for genetic selection. *J. Dairy Sci.* 91:2501–2511.
- de Haas, Y., R. F. Veerkamp, H. W. Barkema, Y. T. Gröhn, and Y. H. Schukken. 2004. Associations between pathogen-specific cases of clinical mastitis and somatic cell count patterns. *J. Dairy Sci.* 87:95–105.
- de Vries, M. J., and R. F. Veerkamp. 2000. Energy balance of dairy cattle in relation to milk production variables and fertility. *J. Dairy Sci.* 83:62–69.
- Detilleux, J., and P. L. Leroy. 2000. Application of a mixed normal mixture model for the estimation of mastitis-related parameters. *J. Dairy Sci.* 83:2341–2349.
- Eding, H., Y. de Haas, and G. de Jong. 2009. Predicting mastitis resistance breeding values from somatic cell count indicator traits. Proc. 2009 Interbull Mtg., Barcelona, Spain. Interbull, Uppsala, Sweden.
- Emanuelson, U., B. Danell, and J. Philipsson. 1988. Genetic parameters for clinical mastitis, somatic cell counts and milk production estimated by multiple-trait restricted maximum likelihood. *J. Dairy Sci.* 71:467–476.
- Gianola, D., and J. L. Foulley. 1983. Sire evaluation for ordered categorical data with a threshold model. *Genet. Sel. Evol.* 15:201–224.
- Green, M. J., L. E. Green, G. F. Medley, Y. H. Schukken, and A. J. Bradley. 2002. Influence of dry period bacterial intramammary infection on clinical mastitis in dairy cows. *J. Dairy Sci.* 85:2589–2599.
- Green, M. J., L. E. Green, Y. H. Schukken, A. J. Bradley, E. J. Peeler, H. W. Barkema, Y. de Haas, V. J. Collis, and G. F. Medley. 2004. Somatic cell count distributions during lactation predict clinical mastitis. *J. Dairy Sci.* 87:1256–1264.
- Hagnestam, C., U. Emanuelson, and B. Berglund. 2007. Yield losses associated with clinical mastitis occurring in different weeks of lactation. *J. Dairy Sci.* 90:2260–2270.
- Harmon, R. J. 1994. Physiology of mastitis and factors affecting somatic cell counts. *J. Dairy Sci.* 77:2103–2112.
- Heringstad, B., G. Klemetsdal, and J. Ruane. 2000. Selection for mastitis resistance in dairy cattle: A review with focus on the situation in the Nordic countries. *Livest. Prod. Sci.* 64:95–106.
- Heringstad, B., R. Rekaya, D. Gianola, G. Klemetsdal, and K. A. Weigel. 2003. Genetic change for clinical mastitis in Norwegian cattle: A threshold model analysis. *J. Dairy Sci.* 86:369–375.
- Heuven, H. C. M. 1987. Diagnostic and genetic analysis of mastitis field data. PhD Thesis. University of Wisconsin, Madison.
- Ibáñez, M. A., M. J. Carabaño, J. L. Foulley, and R. Alenda. 1996. Heterogeneity of herd-period phenotypic variances in the Spanish Holstein-Friesian cattle: Sources of heterogeneity and genetic evaluation. *Livest. Prod. Sci.* 45:137–147.
- Leclerc, H., W. F. Fikse, and V. Ducrocq. 2005. Principal components and factorial approaches for estimating genetic correlations in international sire evaluation. *J. Dairy Sci.* 88:3306–3315.
- Lund, M. S., J. Jensen, and P. H. Petersen. 1999. Estimation of genetic and phenotypic parameters for clinical mastitis, somatic cell production deviance, and protein yield in dairy cattle using Gibbs sampling. *J. Dairy Sci.* 82:1045–1051.
- Madsen, P., M. M. Shariati, and P. Ødegård. 2008. Genetic analysis of somatic cell score in danish holsteins using a liability-normal mixture model. *J. Dairy Sci.* 91:4355–4364.

- Mark, T., W. F. Fikse, U. Emanuelson, and J. Philipsson. 2002. International genetic evaluations of holstein sires for milk somatic cell count and clinical mastitis. *J. Dairy Sci.* 85:2384–2392.
- Miglior, F., B. L. Muir, and B. J. Van Doormaal. 2005. Selection indices in Holstein cattle of various countries. *J. Dairy Sci.* 88:1255–1263.
- Misztal, I., S. Tsuruta, T. Strabel, B. Auvray, T. Druet, and D. H. Lee. 2002. BLUPF90 and related programs (BGF90). Proc. 7th WCGALP, Montpellier, France. CD-ROM communication 28:07.
- Nielsen, C. 2009. Economic impact of mastitis in dairy cows. PhD Thesis. Swedish University of Agricultural Sciences, Uppsala.
- Ødegård, J., J. Jensen, P. Madsen, D. Gianola, G. Klemetsdal, 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.
- Pantoja, J. C. F., P. L. Hullan, and P. L. Ruegg. 2009. Dynamics of somatic cell counts and intramammary infections across the dry period. *Prev. Vet. Med.* 90:43–54.
- Reents, R., J. Jamrozik, L. R. Schaeffer, and J. C. M. Dekkers. 1995. Estimation of genetic parameters for test day records of somatic cell score. *J. Dairy Sci.* 78:2847–2857.
- Soyeurt, H., I. Misztal, and N. Gengler. 2010. Genetic variability of milk components based on mid-infrared spectral data. *J. Dairy Sci.* 93:1722–1728.
- Stoop, W. M., H. Bovenhuis, J. M. L. Heck, and J. A. M. van Arendonk. 2009. Effect of lactation stage and energy status on milk fat composition of Holstein-Friesian cows. *J. Dairy Sci.* 92:1469–1478.
- Vallimont, J. E., C. D. Dechow, C. G. Sattler, and J. S. Clay. 2009. Heritability estimates associated with alternative definitions of mastitis and correlations with somatic cell score and yield. *J. Dairy Sci.* 92:3402–3410.
- Waage, S., S. Sviland, and S. A. Ødegaard. 1998. Identification of risk factors for clinical mastitis in dairy heifers. *J. Dairy Sci.* 81:1275–1284.
- Weigel, K. A., and R. Rekaya. 2000. A multiple-trait herd cluster model for international dairy sire evaluation. *J. Dairy Sci.* 83:815–821.
- Windig, J. J., M. P. L. Calus, G. de Jong, and R. F. Veerkamp. 2005. The association between somatic cell count patterns and milk production prior to mastitis. *Livest. Prod. Sci.* 96:291–299.
- Windig, J. J., W. Ouweltjes, J. ten Napel, G. De Jong, R. F. Veerkamp, and Y. de Haas. 2010. Combining somatic cell count traits for optimal selection against mastitis. *J. Dairy Sci.* 93:1690–1701.
- Zwald, N. R., K. A. Weigel, W. F. Fikse, and R. Rekaya. 2003. Application of a multiple-trait herd cluster model for genetic evaluation of dairy sires from seventeen countries. *J. Dairy Sci.* 86:376–382.