Combining somatic cell count traits for optimal selection against mastitis

J. J. Windig,^{*1} W. Ouweltjes,^{*} J. ten Napel,^{*} G. de Jong,[†] R. F. Veerkamp,^{*} and Y. De Haas^{*†}

^{*}Animal Breeding and Genomics Centre–Livestock Research, Wageningen UR, PO Box 65, NL-8200 AB Lelystad, the Netherlands

[†]CRV, Animal Evaluation Unit, PO Box 454, NL-6800 AL Arnhem, the Netherlands

ABSTRACT

Test-day records of somatic cell counts (SCC) can be used to define alternative traits to decrease genetic susceptibility to clinical mastitis (CM) and subclinical mastitis (SCM). This paper examines which combination of alternative SCC traits can be used best to reduce both CM and SCM and whether direct information on CM is useful in this respect. Genetic correlations between 10 SCC traits and CM and SCM were estimated from 3 independent data sets. The SCC traits with the strongest correlations with CM differed from those with the strongest correlations with SCM. Selection index calculations were made for a breeding goal of 50% CM and 50% SCM resistance using these correlations. They indicated that a combination of 5 SCC traits (SCC early and late in lactation, suspicion of infection based on increased SCC, extent of increased SCC, and presence of a peak pattern in SCC) gave a high accuracy, almost without loss, compared with the full set of 10 SCC traits. The estimated accuracy of this index was 0.91, assuming that the correlations had been estimated without error. To take errors in estimation into account, correlations were resampled from a normal distribution with mean and standard errors as originally estimated. The accuracy of the index calculated with the original correlations was then recalculated using the resampled correlations. The average accuracy based on 50,000 resamplings decreased to 0.81. Use of direct information on CM improved the accuracy (uncorrected for errors in correlations) only slightly, to 0.92.

Key words: somatic cell count, udder health index, mastitis, trait selection

INTRODUCTION

Next to production traits, health and fertility are important in current selection programs. Mastitis is one of the most frequent and costly diseases in the dairy industry (Halasa et al., 2007; Huijps et al., 2008). Besides

management, breeding is an important tool to decrease mastitis incidence. The advantage of reducing mastitis by breeding is that it results in a permanent change in the genetic composition of the dairy herd (Shook, 1989). A decrease in lactation-average SCC is routinely used as a predictor to reduce mastitis incidence. Recently, other SCC-derived traits have been proposed as alternatives for lactation average, to improve breeding for udder health (de Haas et al., 2008).

Several reasons explain why a simple lactation-average SCC is not the optimal way to analyze udder health for breeding purposes. Lactation-average SCC does not distinguish potential differences in origins of increases in SCC. Although IMI explain most of the increases in SCC, other factors, like heat stress, stage of lactation, or the hormonal cycle of the cow, also influence SCC (Wever and Emanuelson, 1989). Lactation-average SCC does not distinguish between lactations with, for example, medium SCC across the lactation and a lactation with single test-day with increased SCC. Moreover, IMI themselves may have different origins (i.e., they are caused by different types of pathogens). From a genetic viewpoint, the consequence can be that SCC is better described as 2 or more traits. Weakly or negatively correlated breeding values between 2 traits may cause low heritabilities for their average if phenotypic correlations are strong and positive. Consequently, the ability to select animals with a high resistance to increased SCC can be reduced when using only lactation-average SCC.

A whole range of different formulations of SCC have been proposed. Here we limit to 3 groups of alternative SCC traits: SCC traits defined on the basis of lactation stage (Haile-Mariam et al., 2001), occurrence of excessive SCC (ten Napel et al., 2009b), and SCC traits on the basis of patterns in peaks of SCC (de Haas et al., 2002b). Increases in SCC early in lactation will include increases caused by IMI from which animals can quickly recover, whereas sustained infections will be reflected more in late-lactation SCC. Somatic cell counts above a certain threshold will be more likely caused by IMI, whereas variation below the threshold will more likely stem from variation within healthy or recovered cows. Finally, patterns of peaks may reflect

Received January 26, 2009.

Accepted December 15, 2009.

¹Corresponding author: Jack.Windig@wur.nl

the type of pathogen causing an IMI, distinguishing between short, intensive infections and long infections with moderate increases of SCC.

Mastitis can be divided into clinical mastitis (CM) and subclinical mastitis (SCM). Clinical mastitis is defined as mastitis with clearly visible symptoms; in practice, these cases are diagnosed by the veterinarian or farmer. Subclinical mastitis cases are those cases that do not lead to visual signs; in practice, they increase SCC but are not diagnosed as IMI. de Haas et al. (2008) analyzed genetic correlations of different SCC traits with CM and SCM. They concluded that some traits were more strongly related to CM and others more strongly related to SCM. Therefore, a combination of different traits would be more effective if both types of mastitis need to be reduced by breeding. Which combination can be used best also depends on the genetic correlations between the alternative SCC traits themselves. Two highly correlated traits are less likely to cover different genetic aspects of udder health than 2 weakly correlated traits.

Frequencies of mastitis differ between parities (Pösö and Mäntysaari, 1996; Sargeant et al., 1998). The resistance against mastitis may also differ depending on whether a cow has experienced an IMI before. Consequently, resistance against mastitis will not be exactly the same trait in different parities (Carlén et al., 2004). The consequence may be that the genetic correlation across parities can be low and that different SCC traits are better used in different parities. On the other hand, if genetic correlations across parities are large, breeding values for resistance against mastitis may be estimated with reasonable accuracy with data available in the first parity only.

Breeding for increased resistance against CM can be performed by direct selection against CM (Heringstad et al., 2003; Carlén et al., 2004) or by indirect selection using traits genetically correlated to CM. In the Netherlands, cases of CM are not recorded, unlike in the Scandinavian countries. Indirect selection may use SCC traits to increase resistance against CM. It is therefore important to know the loss of accuracy of breeding values if no direct information on CM is used in the estimation.

This study examined which combination of alternative traits can be used best to reduce both CM and SCM, and the usefulness of direct information on CM. We first examined the genetic correlation matrices of CM, SCM, and the alternative SCC traits. Next, these correlations were used in selection index calculations to find the optimal combination of SCC traits. To evaluate the usefulness of direct information on CM, the index calculations were repeated with direct information on CM included. Finally, the calculation of genetic correla-

tions was repeated within and across parities to check whether results were stable over parities.

MATERIALS AND METHODS

Trait Definitions

The traits recorded and their means are shown in Table 1. Analyzed udder health traits were CM and SCM, scored as present (1) or absent (0). When at least 1 case of CM or SCM occurred in a lactation, the respective trait was scored as present. No effort was made to distinguish between 1 or more cases because it is often doubtful whether a second case really is a new case or whether it is a continuation of a previous case that was temporarily unnoticed. Cases of SCM were derived from test-day SCC recorded on a monthly basis and defined as a conversion from 2 consecutive SCC recordings below a certain cut-off value to 1 SCC recording above this value (Van den Borne et al., 2006). The cut-off value was 150,000 cells/mL for heifers and 250,000 cells/mL for multiparous cows in this study.

To analyze the additional information from SCC as a mastitis indicator for genetic selection, 3 sets of SCC traits were defined (Table 1; de Haas et al., 2008). In short, the first set of 3 SCC traits consisted of lactation-average SCC over different periods of the lactation, where SCC traits were transformed to SCS using $1,000 + 100 \times [\log(\text{SCC}/1,000,000)]$. The second set of 4 SCC traits was derived from classified test-day SCC recordings (ten Napel et al., 2009b) using threshold levels to distinguish between test-days suspected and unsuspected of IMI. The third set of 3 SCC traits corresponded to patterns of peaks in SCC (de Haas et al., 2002a).

Available Data

Three different data sets with cases of CM were available. The first data set included information from 396 Dutch herds recorded between July 1, 2004 and June 30, 2005. At least 50 milking cows were present on each participating farm. Information on CM was collected by Nederlands Rundvee Syndicaat (NRS, Arnhem, the Netherlands) during the monthly milk recording. The milk samplers asked the farmers whether a cow was mastitic on the day of milk recording or whether she had had a case of CM in the period between the previous and current milk recording. This data set included records for 21,673 cows, with 11,918 first-parity records, 9,991 second-parity records, and 6,779 third-parity records.

The second data set included information from 274 Dutch herds, and data was recorded between July 1,

1999 and January 1, 2006. The actual start and end dates of the data recording varied among farms, depending on when a management information system was purchased and how long it was used on a farm. The farmers themselves collected information on (treatments of) CM, and they allowed NRS to upload and use these data. This data set included records for 30,145 cows, with 23,642 first-parity records, 19,209 second-parity records, and 13,875 third-parity records.

The third data set included information from 795 Dutch herds and was recorded from March 2003 until February 2006. Information on CM was derived from bacteriological testing by assuming that bacteriological testing implied CM. Only the 404 farms that submitted samples for testing at least 5 times in the period of recording were scored for CM; otherwise, CM was set to missing. Farms that tested all cows for CM were excluded because these farms presumably also tested healthy cows. For the 404 farms, CM was scored as present for a cow if a sample was submitted for bacteriological testing and otherwise as not present. All submissions were taken into account, whether by a veterinarian (most cases) or another person. This data set included records for 39,769 cows, with 23,642 first-parity records, 19,209 second-parity records, and 13,875 third-parity records.

The definition of CM differed among the 3 data sets. Genetic correlations between CM as defined in the 3 data sets could be estimated because the sires of the cows in all 3 data sets overlapped. The genetic correlations were 0.94, 0.86, and 0.84, indicating that they were similar traits (Ouweltjes et al., 2008). The structure of the genetic correlation matrix with SCC traits was also very similar across data sets. Therefore, the average of the genetic correlation matrices estimated for the 3 data sets was used for selection index calculations.

Genetic Correlations

(Co)variances were first estimated for the 3 data sets, with observations in different parities regarded as repeated observations on the same trait. Results have been reported previously (de Haas et al., 2008; Ouweltjes et al., 2008) and will not be described extensively here. The model used for estimation was the same as the model used in this paper except that animals could have more observations per trait (i.e., 1 observation/lactation). In this study, observations within a lactation were regarded as a separate trait, so the original traits were split into 3 separate traits, 1 for each of the parities separately. This procedure allowed us to estimate the effect of parities separately, which is needed in practical situations where available data is a mixture of cows with single and multiple lactations.

Table 1. Somatic cell count traits used in this study with their explanations, and means for the 3 data sets analyzed in this study

Trait	Definition	Data set ¹		
		1	2	3
CM	Absence or presence of clinical mastitis (0/1)	0.109	0.104	0.087
SCM	Absence or presence of subclinical mastitis (0/1)	0.351	0.328	0.318
SCS_total	Log-transformation of average test-day SCC up to 400 DIM	667	660	663
SCS_early	Log-transformation of average test-day SCC up to 150 DIM	609	601	619
SCS_late	Log-transformation of average test-day SCC between 151 and 400 DIM	657	646	672
Suspect	Absence or presence of test-day somatic cell counts >150,000 cells/mL (0/1)	0.598	0.571	0.548
Extent	Proportion of test-day SCC >150,000 cells/mL of all recorded test-days	0.394	0.212	0.235
Severe	Absence/presence of test-day SCC >250,000 cells/mL within suspect group (0/1)	0.680	0.678	0.683
Length	Number of consecutive test-days during the longest streak with test-day SCC >150,000 cells/mL within suspected group	1.841	1.568	1.803
P_env	Number of environmental peaks (i.e., low-high-low; high = >150,000 cells/mL)	0.264	0.224	0.266
P_cont	Number of contagious peaks (i.e., low-high-high)	0.130	0.113	0.132
P_all	Sum of environmental and contagious peaks	0.134	0.111	0.134

¹Data set 1: clinical mastitis (CM) recorded by milk samplers; data set 2: CM recorded in management information system of farmer; data set 3: CM recorded by testing of bacteriological sample.

While constructing the data sets, the aim was to include the maximum feasible number of lactations recorded during the study and to avoid bias caused by culling in early lactation. Therefore, every lactation that was recorded for at least 1 d during the experiment was included in the data set. Consequently, in the analyses, traits were corrected for 1) the number of days at risk (number of days of the lactation that fell within the registration period) and DIM for each lactation, and 2) the number of DIM at the start of the registration period, by including them as effects in the model.

The complete history of the milk production recordings (**MPR**) from all cows in all data sets was available. A record included a cow identification number, herd number, date of calving, parity, date of MPR, MPR yields (kg of milk, fat, and protein), and SCC (cells/mL). A pedigree file of all participating cows was also available and included the ancestry of approximately 140,000 animals ($\pm 124,000$ cows and $\pm 16,000$ bulls) back to 1908 and included information on the date of birth and the breed. The number of sires with progeny was 1,903, 3,126, and 3,366 for data sets 1, 2, and 3, respectively. The number of common sires was 1,178 between data sets 1 and 2, 1,242 between data sets 1 and 3, and 1,838 between data sets 2 and 3.

Data editing was done by excluding lactations 1) of fourth parity and higher, 2) with ages at calving < 640 d, and 3) of cows with $< 75\%$ Holstein-Friesian genes. Only test-day SCC recordings between 5 and 400 DIM were included, and only observations on CM between -15 and 400 DIM were included. For the analyses, pedigree files were constructed based on sires and maternal grandsires of cows in the data. Cows with unknown pedigree and cows with sires with < 5 offspring (as sire or maternal grandsire) in the data set were deleted.

Variance components were estimated using generalized linear mixed models with restricted error maximum likelihood via the ASREML package (Gilmour et al., 2006). Bivariate analyses were carried out to estimate correlations between CM, SCM, and SCC traits using a linear model for all traits. Cows with missing values for SCM or SCC traits were still included in the analyses. The model included random effects for sire and maternal grand sire. The model used was

$$y = \mu + h \times y + m + f(\text{ca}) + s_{\text{sire}} + 1/2 s_{\text{MGS}} + e,$$

where y is the trait analyzed, μ is the mean, $h \times y$ is the interaction between herd and year of calving, m is the month of calving (with 12 classes), and $f(\text{ca})$ is a linear polynomial for age at calving; both $h \times m$ and $f(\text{ca})$ are fixed effects; and e is the residual effect. The

random sire effect was identified by the subscripts for sire and MGS (s_{sire} and s_{MGS} , respectively) and, therefore, each animal in the pedigree obtained 1 estimate for the additive genetic effect. The $1/2$ was included because the incidence matrix of MGS effects was laid over the matrix of sire effects (i.e., if a bull had both entries in the data as sire and MGS, the breeding value as MGS was equal to half the breeding value as a sire). Sire effects were linked using the relationship matrix of the sires, and these effects were assumed to be normally distributed with $\text{Var}(s_{\text{sire}}) = \text{Var}(s_{\text{MGS}}) = \sigma_s^2$.

Genetic parameters were calculated from the estimated (co)variance components in the bivariate analyses. If more than 1 estimate was available for a particular variance, the mean was used. The additive genetic variance was calculated as $\sigma_A^2 = 4 \times \sigma_s^2$. The phenotypic variance was calculated as $\sigma_P^2 = 1.25 \times \sigma_s^2 + \sigma_r^2$, with σ_r^2 being the residual variance. The factor 1.25 is explained by the fact that effects for both sires ($1 \times \sigma_s^2$) and MGS ($0.25 \times \sigma_s^2$) explained part of the total genetic variance. Heritabilities were estimated as $h^2 = \sigma_A^2 / \sigma_P^2$.

Linear models were used for all traits. Combined linear and logistic models were also analyzed in which CM, SCM, and the bivariate SCC traits were treated as categorical traits and the other traits were assumed to be normally distributed. The estimated genetic (co) variances and correlations were similar to those originating from the complete linear model. Therefore, only the estimated parameters from the complete linear model are shown. de Haas et al. (2008) also carried out univariate analyses with a logistic model to account for the binary nature of CM and SCM, but conclusions were the same as for linear models and will not be presented here.

Selection Index Calculations

To determine the optimal combination of SCC traits for increasing mastitis resistance, selection index theory was used (Hazel, 1943; Falconer and Mackay, 1996). The selection goal (H) was defined by CM and SCM with equal weights. Different indices (I) were evaluated by calculating their accuracy ($r_{H,I}$). Index calculations were first performed using heritabilities and genetic correlation matrices from the analyses, with the observations in different parities regarded as the same trait. For these calculations, the average of the heritabilities and correlation matrices of the 3 data sets was used. Genetic (co)variance matrices were calculated from the genetic correlations, heritabilities, and phenotypic variances. Two selection indices were used as benchmarks for comparisons with different combinations of

SCC traits. The first benchmark was a selection index with only direct observations on SCM (i.e., a minimal number of observed traits). The second benchmark was an index with the maximum number of traits possible. This index consisted of a combination of SCM with all SCC traits except SCS_total and P_all. These last 2 traits were the sum of 2 other traits (P_env + P_cont and SCS_early + SCS_late, respectively) and combining them with the consisting parts would not have been meaningful.

To select the best combination of traits, the accuracy ($r_{H,I}$) of different combinations were calculated. Because the number of possible combinations of the 10 traits was too large to analyze, a stepwise approach was used. First, each trait was compared separately with the minimum and maximum benchmark. Second, combinations within the 3 sets of traits (e.g., SCC in different periods, SCC derived from threshold values, and SCC peak patterns) were analyzed. Finally, promising combinations from the within-set analyses were combined to find an optimum index.

Calculations were performed using software of Van der Werf (2006). This program uses phenotypic variances, heritabilities, and the genetic correlation matrix as input, constructs the genetic (co)variance matrix from these, and bends it to be positive definite. Phenotypic variances were set to 0 by us so that the weights for the different traits could easily be compared. Accuracies are independent of the phenotypic variances. The accuracy was evaluated by assuming a bull with observations for each trait except CM on 120 offspring. The dams of the offspring were assumed to be a random sample of the population (i.e., their average relatedness was assumed to be the same as that of the whole population). For CM, no observations were assumed. As a consequence, breeding value estimation of CM was assumed to be performed using only observations on the other traits through their genetic correlations with CM (but see below under Use of Direct CM Observations).

With use of multiple traits, the estimated accuracy is an overestimation (Sales and Hill, 1976). Weights $\mathbf{b}$ are estimated based on the estimated genetic and phenotypic (co)variance matrices ($\mathbf{G}$ and $\mathbf{P}$, respectively) so that $r_{H,I}$ is maximized. However, this maximization will be realized only if $\mathbf{G}$ and $\mathbf{P}$ are estimated without error. In practice, $\mathbf{G}$ and $\mathbf{P}$ are not estimated without error and consequently weights $\mathbf{b}$ will not be optimal and the accuracy will be smaller than estimated. Provided that the error in estimating $\mathbf{G}$ and $\mathbf{P}$ is roughly equal across correlations, the ranking of different indices will remain the same. However, the accuracy of the index itself will deviate. Therefore, we reestimated the accuracy of the index that was finally selected. To do this, we simulated new $\mathbf{G}$ and $\mathbf{P}$, denoted $\mathbf{G}^*$ and $\mathbf{P}^*$, by drawing each

individual correlation from a normal distribution with mean and standard deviation given by the original estimate. The accuracy was then estimated by combining the weights calculated for the original $\mathbf{G}$ and $\mathbf{P}$ with the reestimated $\mathbf{G}^*$ and $\mathbf{P}^*$ instead. This procedure was repeated 50,000 times and the final estimation of the accuracy was the mean of the 50,000 estimates.

Use of Direct CM Observations

To evaluate the effect of direct observations of CM on the accuracy of breeding value estimation, the index calculations were also performed assuming that 30, 60, or 100% of the offspring had information on CM. Calculations were performed for the 2 benchmark combinations (only SCM information and the full set) and for 3 promising combinations of SCM and some SCC traits. Accuracy was estimated relative to an index without direct observations on CM.

Selection Index and Different Parities

To evaluate the effect of parity, selection index calculations were repeated using the estimates for heritabilities and genetic correlations, with observations in different parities regarded as separate traits. The average of both data sets used for the estimation was used, except when one of the estimates had not converged and was fixed to a correlation of 1. This occurred for 8 traits in the second data set and 30 traits in the third data set. In that case, only the estimate from the other data set was used. The cause of nonconvergence was probably correlations close or equal to unity in combination with low heritability traits, with many observations being 0.

For 6 genetic correlations, the estimates differed more than 0.25 between the 2 data sets. For 4 of these, the mean of the 2 data sets was used. For 2 of these genetic correlations, the estimate of one of the data sets was chosen. The estimate between SCS_early and extent in the first parity was 0.917 in the second data set and 0.624 in the third data set. In both data sets, genetic correlations of both extent and SCS_early with SCS_late and with suspect were all >0.94 . Consequently, the genetic correlation of 0.624 was unlikely. In this case, the genetic correlation of 0.917 was chosen as the final estimate. Likewise, the correlation in the first data set between SCS_late in the first parity and suspect in the third parity was chosen rather than their mean from the data sets. This decision gave an estimated genetic (co)variance matrix that needed considerably less bending to be positive definite. After bending to obtain a positive definite matrix, the difference with the original matrix was at most 0.06, compared with

Table 2. Genetic correlations for alternative SCC traits¹

Item	CM	SCM	SCS_total	SCS_early	SCS_late	Suspect	Extent	Severe	Length	P_all	P_env
SCM	0.646										
SCS_total	0.796	0.904									
SCS_early	0.820	0.743	— ²								
SCS_late	0.697	0.963	—	0.869							
Suspect	0.738	0.966	0.897	0.856	0.993						
Extent	0.723	0.851	0.971	0.940	0.961	0.942					
Severe	0.797	0.729	0.907	0.895	0.853	0.841	0.930				
Length	0.728	0.886	0.938	0.877	0.953	0.936	0.979	0.815			
P_all	0.879	0.729	0.888	0.949	0.764	0.799	0.909	0.922	0.855		
P_env	0.857	0.698	0.910	0.825	0.727	0.706	0.863	0.889	0.774	—	
P_cont	0.851	0.701	0.778	0.927	0.835	0.846	0.924	0.925	0.905	—	0.885

¹See Table 1 for trait definitions. Mean of 3 data sets (de Haas et al., 2008; Ouweltjes et al., 2008). Standard errors on average = 0.052 (range = 0.003–0.12).

²Dependent variables; correlation not calculated.

0.14 after bending of the matrix where the inconsistent correlations were not ignored.

The 3 parities were first analyzed separately to see whether different combinations of traits were needed for efficient indices in different parities. Finally, the selected combination was tested with traits in different parities considered as different traits. Thus, the selection goal was $(CM_1 + CM_2 + CM_3 + SCM_1 + SCM_2 + SCM_3)/6$, where the number indicates the parity and weights were estimated separately for the observed traits within parities.

RESULTS

Genetic Correlations

Genetic correlations ranged from 0.65 to 0.99 (Table 2). Strong correlations were found between suspect, length, extent, and SCS_late and between the 3 peak variables and SCS_early. The trait SCM was more strongly related to the suspect, length, extent, and SCS_late, and CM to the 3 peak variables and SCS_early. Variables with relatively strong correlations to SCM had relatively weak correlations with CM. The strongest correlation of all the SCC variables with SCM was found for SCS_late. This variable, however, had the weakest correlation with CM of all the SCC variables. The traits CM and SCM had a relatively weak correlation of 0.65.

Selection Index

The accuracy of a selection index for SCM and CM, based only on observed data of SCM (minimal set), was 0.644 (Table 3). An index based on all variables except the dependent variables SCS_total and P_all (full set) had an accuracy of 0.909, an increase of more than

41%. When single variables were added to the minimal set, the largest increases were observed for SCS_total, SCS_early, SCS_late, and extent, 27 to 34% above the minimal set. Severe and P_cont added least to SCM only (7 and 11%, respectively). Removal of a single trait from the full set reduced the accuracy by no more than 1.4% (for SCS_late), indicating that information present in a single trait was generally also present in other traits.

Adding the combination of SCS_early and SCS_late increased the accuracy an extra 2.4% compared with SCS_total (Table 4). Similarly, for peaks, the combination of P_env and P_cont increased the accuracy by an extra 3.2% compared with P_all. The combination of the 2 threshold SCC traits that increased the accuracy most was suspect and extent. Adding a third trait improved the accuracy by at most 0.9%. Combining traits from the 3 sets of traits (i.e., SCC in different periods, SCC derived from threshold levels, and SCC peak patterns) further increased the accuracy. Using a combination of SCS_early, SCS_late, suspect, extent, and P_all resulted in an accuracy of 0.905, close to the accuracy of 0.909 reached by using the full set of traits. The combination of SCS_early, SCS_late, suspect, extent, P_env, and P_cont scarcely improved over the combination of SCS_early, SCS_late, suspect, extent, and P_all, so this latter combination was selected for further investigation.

Standard errors of the estimated correlations ranged from 0.03 to 0.12. When the correlation matrices were resampled to take the error in their estimation into account, the accuracy of the index decreased in 89.2% of the cases. The accuracy after resampling ranged from 0.535 to 1.000, with an average of 0.814. The distribution was skewed to the left because accuracies >1.0 are impossible. Nevertheless, the median of 0.813 was very similar to the average. The standard deviation of the simulated accuracies was 0.069.

Table 3. Accuracy of indices with subclinical mastitis (SCM) plus an alternative trait or of the full set without the same trait¹

Trait (SCM +)	Trait added to SCM only			Trait removed from full set		
	Accuracy	Relative to		Accuracy	Relative to	
		Minimum (%)	Maximum (%)		Minimum (%)	Maximum (%)
No extra trait ²	0.644	0.0	−29.1	0.899	39.6	−1.1
SCS_total	0.864	34.1	−5.0			
SCS_early	0.819	27.1	−9.9	0.903	40.1	−0.7
SCS_late	0.845	31.1	−7.1	0.896	39.0	−1.4
Suspect	0.807	25.3	−11.2	0.900	39.6	−1.0
Extent	0.811	25.8	−10.8	0.908	40.9	−0.1
Severe	0.683	6.1	−24.8	0.909	41.0	0.0
Length	0.722	12.1	−20.6	0.908	40.9	−0.1
P_all	0.789	21.2	−14.1			
P_env	0.778	20.7	−14.5	0.898	39.3	−1.3
P_cont	0.720	11.7	−20.8	0.908	40.9	−0.1
Full set ³	0.909	41.1	0.0			

¹See Table 1 for trait definitions. Accuracy is estimated for a bull with 120 daughters. Relative accuracy is $\text{accuracy}(\text{set})/\text{accuracy}(\text{minimum}) - 1$ or $1 - \text{accuracy}(\text{set})/\text{accuracy}(\text{maximum})$, where $\text{accuracy}(\text{minimum})$ is the accuracy of an index with only SCM, and $\text{accuracy}(\text{maximum})$ is the accuracy of the full set.

²With $\text{accuracy}(\text{minimum})$.

³Full set with $\text{accuracy}(\text{maximum})$: SCM, SCS_early, SCS_late, suspect, extent, severe, length, P_env, P_cont.

Use of Direct CM Observations

The use of direct information on CM increased the accuracy. The magnitude of the increase depended on

how many other SCC traits the index included and how many offspring had information on CM (Table 5). When only SCM was used in the index, adding CM information for all offspring increased the accuracy

Table 4. Accuracy of index with different combinations of alternative SCC traits¹

Trait(s) added to SCM only	Accuracy	Relative to	
		Minimum (%)	Maximum (%)
SCS_total	0.864	34.1	−5.0
SCS_early, SCS_late	0.880	36.5	−3.2
Suspect, extent	0.856	32.8	−5.9
Suspect, severe	0.818	26.9	−10.1
Extent, severe	0.816	26.6	−10.2
Suspect, extent, severe	0.859	33.3	−5.5
Suspect, length	0.827	28.4	−9.0
Extent, length	0.823	27.7	−9.5
Suspect, extent, length	0.862	33.7	−5.2
Severe, length	0.745	15.7	−18.0
Suspect, severe, length	0.835	29.6	−8.1
Extent, severe, length	0.828	28.4	−9.0
Suspect, extent, severe, length	0.864	34.1	−4.9
P_all	0.789	21.2	−14.1
P_cont, P_env	0.801	24.4	−11.8
SCS_total, suspect, extent	0.891	38.3	−1.9
SCS_early, SCS_late, suspect, extent	0.899	39.5	−1.1
SCS_total, suspect, extent, P_all	0.892	38.4	−1.9
SCS_early, SCS_late, suspect, extent, P_all	0.905	40.5	−0.4
SCS_early, SCS_late, suspect, extent, P_env, P_cont	0.907	40.8	−0.2
Full set	0.909		

¹See Table 1 for trait definitions. Accuracy is estimated for a bull with 120 daughters. Relative accuracy is $\text{accuracy}(\text{set})/\text{accuracy}(\text{minimum}) - 1$ or $1 - \text{accuracy}(\text{set})/\text{accuracy}(\text{maximum})$, where $\text{accuracy}(\text{minimum})$ is the accuracy of an index with only subclinical mastitis (SCM) and $\text{accuracy}(\text{maximum})$ is the accuracy of the full set. The combination of traits selected as most efficient is in bold.

Table 5. Effect of adding direct information on clinical mastitis (CM) on accuracy of selection indices¹

Item	Accuracy				Gain vs. no CM data (%)		
	Offspring with data on CM (n)				Offspring with data on CM (n)		
	0	30	60	120	30	60	120
Only (S)CM	0.644	0.689	0.722	0.767	7.0	12.1	24.1
SCS_early, SCS_late	0.880	0.887	0.893	0.903	3.7	4.5	5.6
SCS_early, SCS_late, suspect, extent	0.899	0.905	0.910	0.918	0.7	1.2	2.1
SCS_early, SCS_late, suspect, extent, P_all	0.905	0.913	0.916	0.922	0.5	1.0	1.7
Full set	0.909	0.913	0.917	0.923	0.5	0.9	1.6

¹See Table 1 for trait definitions. Accuracy is estimated for a bull with 120 daughters on all traits except CM. For CM, information of daughters was varied from 0 to all 120 daughters.

18.8%. The increase in accuracy decreased to 1.4% if CM was added to the full set of SCC traits, and decreased to 1.5% if CM was added to the set with SCS_early, SCS_late, suspect, extent, and P_all. When information on CM was available for only a part of the offspring, the increase in accuracy was less.

Different Parities

Genetic correlations within parities (Table 6) were similar to correlations calculated over all parities (Table 2). Genetic correlations between CM and SCM were relatively weak, especially in third parity (0.26). Correlations between SCM and SCC traits tended to be larger (range = 0.63–0.99) than between CM and SCC traits (range = 0.51–0.87). In first parity, correlations between CM and SCC traits ranged from 0.70 to 0.78, whereas in the second and third parities they ranged from 0.53 to 0.87. The SCC traits with a relatively strong correlation with SCM tended to have a relatively weak correlation with CM. Correlations of SCC traits with other SCC traits were often strong (i.e., 16 correlations >0.95), indicating that they include similar information. Correlations among SCC traits ranged from 0.68 to 0.99 within the same period and from 0.54 to 0.99 across parties. The weakest correlations across parities were found between parities 1 and 3, and the strongest correlations between parities 2 and 3. Interestingly, the correlations between CM or SCM across parities were sometimes smaller than between CM or SCM and another SCC trait across parities. For example, the correlation between CM in the third and first parities was 0.67, whereas the correlation between CM in the third parity and P_all in the first parity was 0.79. Likewise, the correlation between SCM in the first and third parities was smaller than between SCM in the first parity and all SCC traits in the third parity.

When indices were evaluated separately for parities, results were similar to an analysis without information on parities (Table 7). Weights for the observational

traits for the index with observations on SCM, SCS_early, SCS_late, suspect, extent, and P_all followed the same pattern within parities (i.e., largest weights for SCM and SCS_late, smallest weights for extent). Accuracies were 0.92, 0.91, and 0.92 in parities 1, 2, and 3, respectively, which is similar to the accuracy of 0.91 when parity information is ignored.

When a single index was constructed with each trait in a parity treated as a separate trait, the accuracy increased to 0.95 (Table 7). Weights for the traits were different from the within-parity analysis. For SCS_late in the second parity, the weight was 0.00, indicating that its information on SCM and CM was contained in other parities or other SCC traits. The weight for P_all was relatively large across all 3 parities.

DISCUSSION

This research clearly shows that much more information can be gained from SCC test-day records than from the simple lactation-average SCC. The gain in accuracy in estimation of breeding value for mastitis resistance clearly increases when using several alternative SCC traits compared with lactation-average SCC alone. The accuracy is also greater than the estimated accuracy of 0.73 (uncorrected for use of multiple traits) of the current index in the Netherlands, which includes lactation-average SCC, fore udder attachment, udder depth, teat length, and milking speed (De Jong and Lansbergen, 1996). Using a combination of alternative SCC traits eliminates the need to use information from conformation traits. Although recording of CM remains important for herd management by farmers and veterinarians, the use of direct information on CM for breeding value estimation is less urgent when a combination of SCC traits is available. Finally, the use of alternative SCC traits opens up the possibility of including both CM and SCM in the breeding goal.

We choose in this study to concentrate on the improvement of overall udder health. Therefore, we aimed

Table 6. Genetic correlations for selected traits within parities¹

Parity	Trait	h ²	Parity1							Parity 2							Parity 3					
			CM	SCM	SCS early	SCS late	Susp	Extent	P_all	CM	SCM	SCS early	SCS late	Susp	Extent	P_all	CM	SCM	SCS early	SCS late	Susp	Extent
1	CM	0.029																				
1	SCM	0.061	0.578																			
1	SCSearly	0.097	0.749	0.839																		
1	SCSlate	0.126	0.717	0.920	0.889																	
1	Susp	0.072	0.698	0.945	0.969	0.961																
1	Extent	0.114	0.777	0.937	0.901	0.944	0.977															
1	P_all	0.068	0.782	0.768	0.825	0.770	0.906	0.923														
2	CM	0.024	0.687	0.592	0.636	0.582	0.679	0.672	0.872													
2	SCM	0.029	0.548	0.689	0.558	0.768	0.842	0.993	0.681	0.554												
2	SCSearly	0.102	0.663	0.866	0.861	0.857	0.923	0.973	0.854	0.763	0.720											
2	SCSlate	0.123	0.514	0.878	0.705	0.875	0.856	0.805	0.706	0.551	0.898	0.824										
2	Susp	0.071	0.549	0.942	0.715	0.875	0.880	0.789	0.763	0.643	0.993	0.803	0.977									
2	Extent	0.122	0.601	0.956	0.800	0.919	0.928	0.892	0.849	0.762	0.796	0.922	0.968	0.941								
2	P_all	0.051	0.771	0.780	0.813	0.766	0.863	0.893	0.892	0.859	0.599	0.894	0.689	0.691	0.863							
3	CM	0.022	0.666	0.511	0.644	0.500	0.646	0.607	0.793	0.813	0.513	0.770	0.545	0.607	0.687	0.807						
3	SCM	0.025	0.349	0.694	0.439	0.690	0.598	0.874	0.408	0.365	0.809	0.608	0.827	0.874	0.776	0.517	0.259					
3	SCSearly	0.088	0.724	0.756	0.827	0.779	0.838	0.850	0.869	0.822	0.736	0.951	0.785	0.873	0.914	0.916	0.871	0.579				
3	SCSlate	0.120	0.559	0.821	0.684	0.861	0.771	0.743	0.678	0.629	0.895	0.816	0.930	0.987	0.966	0.755	0.535	0.867	0.803			
3	Susp	0.068	0.566	0.904	0.544	0.853	0.727	0.653	0.673	0.597	0.986	0.802	0.805	0.973	0.939	0.706	0.555	0.911	0.824	0.989		
3	Extent	0.096	0.706	0.897	0.784	0.839	0.888	0.888	0.851	0.806	0.946	0.938	0.948	0.985	0.992	0.869	0.760	0.751	0.930	0.676	0.958	
3	P_all	0.068	0.666	0.719	0.728	0.708	0.742	0.764	0.836	0.817	0.732	0.916	0.770	0.824	0.899	0.885	0.846	0.632	0.962	0.772	0.858	0.937

¹See Table 1 for trait definitions. Susp = suspect.

Table 7. Comparison of weights estimated for index on clinical mastitis (CM) and subclinical mastitis (SCM)¹

Trait	Ignored	3 × 6 traits			1 × 18 traits		
		Parity 1	Parity 2	Parity 3	Parity 1	Parity 2	Parity 3
SCM	−0.201	−0.220	−0.214	−0.183	−0.020	−0.046	−0.015
SCS_early	−0.140	−0.148	−0.143	−0.132	−0.050	−0.056	−0.111
SCS_late	−0.218	−0.205	−0.243	−0.188	−0.035	0.000	−0.038
Suspect	−0.220	−0.224	−0.181	−0.240	−0.053	−0.033	−0.050
Extent	−0.089	−0.162	−0.060	−0.106	−0.097	−0.056	−0.111
P_all	−0.205	−0.151	−0.157	−0.158	−0.159	−0.157	−0.088

¹See Table 1 for trait definitions. Ignored: the same trait in different parities is considered as a single trait with repeated observations (i.e., parities are ignored); 3 × 6 traits: parities analyzed separately; 1 × 18 traits: parities analyzed together.

to reduce both CM and SCM, which differs from the general approach toward mastitis concentrating on CM only. The breeding goal was defined as an equal reduction in CM and SCM. We did not use an economic analysis to determine the weights for CM and SCM because, apart from difficulties in determining economic weights, the end goal of the index is the improvement of animal health, not only because of the economic loss attached to poor health but also for improvement of animal welfare. Alternatively, one may emphasize different aspects such as animal welfare in the economic weights (Nielsen et al., 2005). Further refinements may be to weight parities differently; for example, for production traits in the Netherlands, weights of 0.41, 0.33, and 0.26 are used for parities 1, 2, and 3, respectively, mainly based on the frequency of cows in the different parities.

In general, 2 strategies that are not mutually exclusive may be followed in improvement of animal health. The first strategy aims to eliminate pathogens and their spread as much as possible. The second emphasizes improving the resistance and robustness of the animals, including the self-healing capacity of animals (Lam et al., 1997; ten Napel et al., 2009a). In the first strategy, SCM is important because cows that do not show visual signs of mastitis but have an increased SCC may be carriers and spreaders of pathogens infecting other cows. In the second strategy, these cows may be seen as cows that deal with the pathogens themselves, and thus reduction of SCM is unwanted. However, one cannot equate SCM, as defined in this study, completely with self-healing cows. The trait SCM was defined as an increase in SCC and included cases in which an increase in SCC coincided with CM. Moreover, CM was defined here as a case where the farmer diagnosed the cow as mastitic in his management system (second data set) or where the farmer decided to send in a sample for bacteriological testing (third data set). In both cases, some animals that are diseased may be missed

or ignored (Lam et al., 1993). For example, when an animal is already due to be culled farmers will not send in a sample for bacteriological testing. Thus, from an animal health point of view, completely ignoring SCM seems unwarranted.

Different pathogens may challenge different parts of the immune system (Schukken et al., 1997). Consequently, a pathogen-specific breeding program might be more effective than selecting for an overall decrease in CM or SCM. However, resistances to different pathogens interact with each other (Lam et al., 1997), and it would be undesirable to breed for resistance against one pathogen if resistance against another pathogen was subsequently decreased. Consequently, one has to breed for resistance against different pathogens simultaneously. Theoretically, the best way to do this is by constructing an index using observations of cases of CM or SCM that are typed for the causative pathogen. However, it is unlikely that routine typing for all causative pathogens can be realized in the near future in the Netherlands and that the effort matches the increase in effectiveness. Different pathogens influence different SCC traits differently (e.g., Schukken et al., 2003), and most of this variation will be captured in the index described in this study. Consequently, from a breeding perspective it is doubtful that typing all pathogens causing CM or SCM would increase the effectiveness of a breeding program substantially.

The trait SCM was defined as a conversion from 2 consecutive SCC recordings below a certain cut-off value to 1 SCC recording above this value (Van den Borne et al., 2006). This definition was chosen based on veterinary considerations such as incidence rate and length of infection. A consequence is that SCM at the start of the lactation may be missed. Another consequence is that the correlation with SCS_early is automatically smaller. The trait SCS_early is based on the first 150 DIM, whereas SCM cannot occur before the third test-day, which was on average around d 75.

Therefore, the different genetic correlations of SCM and CM with SCS_early can be partly explained by the definition of SCM. However, the different genetic correlation of CM and SCM with SCS_late, peaks, and extent cannot be explained by the definition of SCM.

The trait CM had stronger correlations with SCS_early and peaks, and SCM with SCS_late and extent. This result may indicate that animals are better able to deal with mastitis and, later in lactation, that SCM cases do not become clinical. Alternatively, less severe cases may not be treated early in lactation, with SCC remaining high late in lactation, or cases later in lactation may be recognized less and therefore be reported less as clinical. The consequence of the different correlations with CM and SCM is that the index depends on the (economic) weights given to CM and SCM in the breeding goal. However, the choice of the traits composing the index did not differ when the breeding goal was 100% CM or 100% SCM. The accuracy of an index with the SCC traits selected in this study was 0.88 for 100% CM and 0.94 for 100% SCM (uncorrected for use of multiple traits). When CM receives double the weight of SCM (0.67% vs. 0.33%), as it is probably economically more costly, an accuracy of 0.89 is found.

One should bear in mind that when using multiple traits, the estimated accuracy is an overestimation (Sales and Hill, 1976). This is not a problem as long as one is comparing the accuracies of different indices estimated from genetic (co)variance matrices. However, one may expect a larger accuracy when introducing a new index if it is based on estimates from a genetic (co)variance only. Therefore, a resampling procedure as used in this study is recommended. Here the accuracy decreased from 0.905 to 0.813.

In this study, we used alternative traits derived from SCS to estimate breeding values for CM and SCM combined. Variations in SCS across lactation can also be captured using alternative models such as test-day models (e.g., Madsen et al., 2008; Hagnestam-Nielsen et al., 2009). We used alternative trait definitions because they are biologically related to mastitis infections (de Haas et al., 2002a; ten Napel et al., 2009b). In test-day models, functions are used to describe variation in SCS across the lactation. Some parameters of these functions may be less related to CM and SCM, whereas on the other hand these functions may capture variation related to CM and SCM not captured by our alternative trait definitions. Consequently, both approaches have their own merit and need to be explored.

Using a set of SCC traits can partly overcome absence of direct observations on CM. Direct observations on CM add little to the accuracy of an index if 5 or more SCC traits are used. The consequence is not that direct information on CM is useless, but rather that the

urgency to set up a full infrastructure to collect data on CM is less. The strategy in the Netherlands will be to incorporate direct CM data once they become available through management information systems rather than to set up a full infrastructure such as is already present in the Scandinavian countries.

ACKNOWLEDGMENTS

The farmers are greatly acknowledged for providing information on occurrence of cases of clinical mastitis. This study is part of the 5-yr mastitis program of the Dutch Udder Health Center (Deventer, the Netherlands) and was financially supported by the Dutch Dairy Board (Zoetermeer, the Netherlands). The authors thank Theo Lam and Otlis Sampimon (Dutch Animal Health Service, Deventer, the Netherlands) for their fruitful discussions.

REFERENCES

- 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, and R. F. Veerkamp. 2002a. The effect of pathogen-specific clinical mastitis on the lactation curve of somatic cell count. *J. Dairy Sci.* 85:1314–1323.
- de Haas, Y., H. W. Barkema, and R. F. Veerkamp. 2002b. Genetic parameters of pathogen-specific incidence of clinical mastitis in dairy cows. *Anim. Sci.* 74:233–242.
- 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 Jong, G., and L. Lansbergen. 1996. Udder health index: Selection for mastitis resistance. *Interbull Bull.* 12:42–47.
- Falconer, D. S., and T. F. C. Mackay. 1996. *Introduction to Quantitative Genetics*. 4th ed. Longman Group, Harlow, UK.
- Gilmour, A. R., B. R. Gogel, B. R. Cullis, and R. Thompson. 2006. *ASREML User Guide*. Release 2.0. VSN International Ltd., Hemel Hempstead, UK.
- Hagnestam-Nielsen, C., U. Emanuelson, B. Berghlund, and E. Strandberg. 2009. Relationship between somatic cell count and milk yield in different stages of lactation. *J. Dairy Sci.* 92:3124–3133.
- Haile-Mariam, M., M. E. Goddard, and P. J. Bowman. 2001. Estimates of genetic parameters for daily somatic cell count of Australian dairy cattle. *J. Dairy Sci.* 84:1255–1264.
- Halasa, T., K. Huijps, O. Osteras, and H. Hogeveen. 2007. Economic effects of bovine mastitis and mastitis management: A review. *Vet. Q.* 29:18–31.
- Hazel, L. N. 1943. The genetic basis for constructing selection indexes. *Genetics* 28:476–490.
- 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.
- Huijps, K., T. Lam, and H. Hogeveen. 2008. Costs of mastitis: Facts and perception. *J. Dairy Res.* 75:113–120.
- Lam, T. J. G. M., Y. H. Schukken, F. J. Grommers, J. A. H. Smit, and A. Brand. 1993. Within-herd and between-herd variation in diagnosis of clinical mastitis in cattle. *J. Am. Vet. Med. Assoc.* 202:938–942.
- Lam, T. J. G. M., Y. H. Schukken, J. H. Van Vliet, F. J. Grommers, M. J. M. Tielen, and A. Brand. 1997. Effect of natural infection with minor pathogens on susceptibility to natural infection with

- major pathogens in the bovine mammary gland. *Am. J. Vet. Res.* 58:17–22.
- Madsen, P., M. M. Shariati, and J. Odegard. 2008. Genetic analysis of somatic cell score in Danish Holsteins using a liability-normal mixture model. *J. Dairy Sci.* 91:4355–4364.
- Nielsen, H. M., L. G. Christensen, and A. F. Groen. 2005. Derivation of sustainable breeding goals for dairy cattle using selection index theory. *J. Dairy Sci.* 88:1882–1890.
- Ouweltjes, W., J. J. Windig, G. de Jong, T. J. G. M. Lam, J. ten Napel, and Y. de Haas. 2008. The use of data from sampling for bacteriology for genetic selection against clinical mastitis. *J. Dairy Sci.* 91:4860–4870.
- Pösö, J., and E. A. Mäntysaari. 1996. Relationships between clinical mastitis, somatic cell score, and production for the first three lactations of Finnish Ayrshire. *J. Dairy Sci.* 79:1284–1291.
- Sales, J., and W. G. Hill. 1976. Effect of sampling errors on efficiency of selection indexes. 2. Use of information on associated traits for improvement of a single important trait. *Anim. Prod.* 23:1–14.
- Sargeant, J. M., H. M. Scott, K. E. Leslie, M. J. Ireland, and A. Bashiri. 1998. Clinical mastitis in dairy cattle in Ontario: Frequency of occurrence and bacteriological isolates. *Can. Vet. J.* 39:33–38.
- Schukken, Y. H., T. J. G. M. Lam, and H. W. Barkema. 1997. Biological basis for selection on udder health traits. *Interbull Bull.* 15:27–33.
- Schukken, Y. H., D. J. Wilson, F. Welcome, L. Garrison-Tikofsky, and R. N. Gonzalez. 2003. Monitoring udder health and milk quality using somatic cell counts. *Vet. Res.* 34:579–596.
- Shook, G. E. 1989. Selection for disease resistance. *J. Dairy Sci.* 72:1349–1362.
- ten Napel, J., M. P. L. Calus, H. A. Mulder, and R. F. Veerkamp. 2009a. Genetic concepts to improve robustness of dairy cows. Pages 33–42 in *Breeding for Robustness in Cattle*. M. Klopčič, R. Reents, J. Philipsson, and A. Kuipers, ed. Wageningen Academic Publishers, Wageningen, the Netherlands.
- ten Napel, J., Y. de Haas, G. de Jong, T. J. G. M. Lam, W. Ouweltjes, and J. J. Windig. 2009b. Characterization of distributions of somatic cell counts. *J. Dairy Sci.* 92:1253–1264.
- Van den Borne, B. H. G., G. van Schaik, G. De Jong, and T. J. G. M. Lam. 2006. Determination of a definition of subclinical mastitis incidence based on somatic cell counts in the Netherlands. Pages 609–611 in *11th Int. Symp. Vet. Epidemiol. Econ. (ISVEE)*. ISVEE, Cairns, Australia.
- Van der Werf, J. H. J. 2006. Teaching software used for quantitative genetics at UNE: Multiple trait selection index. <http://www-personal.une.edu.au/~jvanderw/software.htm> Accessed Dec. 12, 2008.
- Wever, P., and U. Emanuelson. 1989. Effects of systematic influences and intrammary infection on differential and total somatic cell counts in quarter milk samples from dairy cows. *Acta Vet. Scand.* 30:465–474.