Predicting Somatic Cell Count Standard Violations Based on Herd's Bulk Tank Somatic Cell Count. Part II: Consistency Index

J. M. Lukas,* J. K. Reneau,*¹ C. Munoz-Zanzi,† and M. L. Kinsel‡

*Department of Animal Science, and

†School of Public Health, University of Minnesota, St. Paul 55108

‡Agricultural Information Management Inc., Ellensburg, WA 98926

ABSTRACT

The present study examines the capability of 1,501 herds in the Upper Midwest and the performance of statistical process control charts and indices as a way of monitoring and controlling milk quality on the farm. For 24 mo, daily or every other day bulk tank somatic cell count (SCC) data were collected. Consistency indices for 5 different SCC standards were developed. The indices calculate the maximum variation allowed to meet a desired SCC level at a given mean bulk tank SCC and were used to identify herds not capable of meeting a specific SCC standard. Consistency index method was compared with a test identifying future bulk tank SCC standard violators based on herds' past violations. The performance of the consistency index test and the past violation method was evaluated by logistic regression. The comparison focused on detection probability and certainty associated with a result. For the 5 SCC levels, detection probability and certainty associated with a result ranged from 51 to 98%. Detection probability of all violators and certainty associated with a negative result was greater for the consistency index across all 5 SCC levels (by 0.7 to 7.4% and 2.1 to 5.1%, respectively). Control charts were plotted and monthly consistency indices calculated for individual farms. Charts in combination with the consistency indices would warn from 66 to 80% of the herds about an upcoming violation within 30 d before it occurred. They offer a proactive approach to maintaining consistently high milk quality. By assessing process capability and distinguishing between significant changes and random variation in bulk tank SCC, tools presented in this article encourage fact-based decisions in dairy farm milk quality management.

Key words: bulk tank somatic cell count, capability index, statistical process control

INTRODUCTION

Statistical process control (SPC), one of the total quality management tools, uses charts to access process performance not only based on its mean but also on the variation in its output. In the last 26 yr numerous attempts have been made to use SPC in animal production (Sard, 1979; Kniffen, 1994; de Vries and Conlin, 2003). Researchers have suggested plotting bulk tank SCC (BTSCC) on SPC charts to monitor and help control milk quality on the farm (Reneau and Kinsel, 2001). Bulk tank SCC serve as an important quality characteristic (Eberhart et al., 1982), constituting the basis for milk quality classification, pricing, and acceptability for global markets. Bulk tank SCC have been also related to management intensity (Barkema et al., 1998) and used as one of the diagnostic tools for milking or bedding routine problems (Chassagne et al., 2005; Reneau et al., 2005). Elevated BTSCC themselves will not pinpoint the source of the milk quality problem; however, they can serve as a signal for further investigation. It is important that farmers and milk processors have signals that help them differentiate between true increase or decrease in BTSCC and random variation that is expected on the farm.

Statistical process control charts are often used to plot current process mean, and variations, based on samples of process output. Current process mean is estimated from the sample average. Process variation, in SPC referred to as sigma, is most commonly estimated from the within-sample standard deviation, unless sample size equals 1. Then process variation is estimated from the average difference between 2 subsequent values, referred to in SPC as average moving range. Statistical process charts focus on separating significant shifts in mean or sigma from random variation observed in the process (Montgomery, 2005).

In manufacturing, to assess the prospect of meeting specification, SPC users calculate capability indices. These indices are commonly calculated by subtracting the mean of process output from the specifications the process is required to meet and dividing by $3 \times$ process sigma (Montgomery, 2005). Depending on the strin-

Received August 28, 2007.

Accepted October 3, 2007.

¹Corresponding author: renea001@umn.edu

gency of the quality control system, a value between 1 and 2 is indicative of a capable process. This concept can easily be adapted in assessment of processes present on the dairy farm. Using legal SCC limits as process specifications, Niza-Ribeiro et al. (2004) have recently introduced capability indices to BTSCC analysis.

The objective of this study was to demonstrate the concept of using variation analysis to assess the prospect of meeting different SCC specifications. When trying to introduce any management tool in an agricultural setting, it is important to make the concept clear and intuitive for the farmer. The language of process capability is not innate to dairy production vocabulary; however, the issue of consistency is easily understood and appreciated. All dairy personnel can relate to the idea that consistent performance of their work logically yields better results.

In our research we attempted to examine the performance of SPC charts and indices when used to help monitor and control milk quality on the farm. To achieve the assumed objective, specific aims were outlined as follows: 1) develop consistency indices (CI) for 5 different SCC standards that would calculate the maximum variation allowed to meet a desired SCC level at a given mean BTSCC; 2) compare the percent correctly identified, detection probability, and certainty associated with a result of a test identifying future SCC standard violators based on herds' current CI or past violations (PV); and 3) study the efficiency of SPC charts and CI as a warning system of future SCC standard violations.

MATERIALS AND METHODS

Study Data Set

Bulk tank SCC data were collected daily or every other day for 24 mo (January 2003 until December 2004), from 1,501 (6.4% of the total population) Upper Midwest (North and South Dakota, Wisconsin, and Minnesota) dairy farms. To be included in the data set, herds had to have BTSCC data from at least 2 subsequent months in 2003 or 2004. Also, information on the average amount of milk shipped monthly had to be available for each farm in the data set. This data was used to divide the herds into 5 herd production categories (HPC) based on the average amount of milk shipped each month in 2003 (USDA classification). Herd production category #1 included all herds with an average amount of milk shipped less than 22,680 kg, #2 included between 22,680 and 45,360 kg, #3 included between 45,360 and 90,720 kg, #4 included between 90,720 and 226,800 kg, and #5 included greater than 226,800 kg. Assuming an average of 30 kg/cow per d, HPC 1, 2, 3, 4, and 5 corresponded to fewer than 25

cows, from 25 to 49 cows, from 50 to 99 cows, from 100 to 249 cows, and 250 or more cows, respectively.

Fifty percent of herds had less than 50 cows (vs. 49% in the United States), 28% had between 50 and 99 head (vs. 30% in the United States), and 22% of herds had 100 or more cows (vs. 21% in the United States). Therefore, it is assumed that the herds in the sample data set had a herd size distribution similar to the one reported in the United States in 2001 (NAHMS, 2003).

Developing the CI

Five different SCC levels (200,000, 300,000, 400,000, 500,000, and 600,000 cells/mL) were used to develop a lookup chart for 5 CI. All 5 CI had the same basic formula: $CI = (\text{standard} - \text{mean})/3$, where the mean is the herd average BTSCC and standard is 1 of the 5 SCC levels. This CI can be interpreted as the maximum variation that may be allowed, given the current mean, for which the process output is statistically assured to stay below the upper specification/standard.

Percent Correctly Identified, Detection Probability, and Certainty Associated with a Result

Because only a single sample is taken from each bulk tank, BTSCC variation (sigma) cannot be estimated from the within sample standard deviation. Sigma was estimated by the moving range of size 2, averaged for each month (AmR), and divided by a constant (sigma = AmR/d_2 , where $d_2 = 1.13$). The monthly means were arithmetic means of individual BTSCC for each herd. Monthly means and sigmas were calculated from individual BTSCC values in 2003 or 2004, for each of the 1,501 herds.

Each month, the maximum BTSCC in each herd was recorded. It was then used to establish the SCC standard violation status of the herd separately, for each of the 5 SCC levels. Herd was declared positive for standard violation (V+) if the maximum BTSCC for that month was greater than the SCC standard, and negative (V-) when the maximum BTSCC was equal or less than the SCC standard.

Consistency Index Method. Bulk tank SCC data from 2 yr were used to evaluate percentage correctly identified, detection probability, and certainty associated with the result of the CI test. Analysis was done for each standard separately. To account for the effect of season on the BTSCC, analysis was performed on a monthly basis. Each month a positive or negative test result was obtained by comparing the CI with the sigma calculated from the BTSCC of that month. A positive CI test (CIT+) was declared when the CI was smaller than the calculated sigma. A negative CI test (CIT-)

		True violation status of the herd	
		Violation of the BTSCC standard in month = j + 1 (V+)	No violation of the BTSCC standard in month = j + 1 (V-)
Result of alternative tests applied in month j	Positive test (CIT+, PVT+)	True positives (TP)	False positives (FP)
	Negative test (CIT-, PVT-)	False negatives (FN)	True negatives (TN)

Figure 1. Grid outlining the procedure of identifying true or false positive or negative results of 2 alternative tests: consistency index test (CI) and past violation test (PV).

was declared when the CI was greater or equal to the calculated sigma. True and false positives and negatives were identified by analyzing the BTSCC data from the following month for the CIT+ and CIT- herds as outlined by the grid in Figure 1.

Past Violation Method. Another screening method, based on past violations (PV), was designed and compared with the performance of the CI. Each month a positive test (PVT+) was declared when the maximum BTSCC for that month exceeded the SCC standard. A negative test (PVT-) was declared when the maximum BTSCC from that month did not exceed the SCC standard. True and false positives and negatives were identified by analyzing the BTSCC data from the current and subsequent month as outlined by the grid in Figure 1.

Analysis. For both methods, percentage correctly identified, PCI_{ijk} , for a herd of HPC i (1 to 5), in month j (1 to 12) and year k (2003 or 2004) was estimated for each SCC standard as

$$PCI_{ijk} = (TP_{ijk} + TN_{ijk}) / (TP_{ijk} + TN_{ijk} + FP_{ijk} + FN_{ijk}),$$

where TP is the number of true positives, TN is the number of true negatives, FP is the number of false positives, and FN is the number of false negatives. Factorial ANOVA (PROC GLM procedure of SAS) was used to evaluate the effect of testing method on percentage correctly identified.

Logistic regression (PROC LOGISTIC of SAS) was used to estimate detection probability and certainty associated with a result for each standard separately. The following factors were considered: year, month,

HPC, and method. To investigate the differences in relationship between detection probability, and certainty associated with a result and the remaining 3 main factors, for the 2 test methods (CI and PV), the 3 interactions of method with year, HPC, and month were also entered into the model. The following represents a general full model used:

$$\begin{aligned} \mu_{ijkl} = & \alpha + \beta_{1i}HPC_i + \beta_{2j}month_j + \beta_{3k}year_k + \beta_{4l}method_l \\ & + \beta_{5il}(HPC_i \times method_l) + \beta_{6jl}(month_j \times method_l) \\ & + \beta_{7kl}(year_k \times method_l). \end{aligned}$$

Subsequently, detection probability and certainty associated with a result was calculated as

$$y_{ijkl} = e^{\mu} / (1 + e^{\mu}),$$

where y_{ijkl} represents the detection probability or certainty associated with a result for a herd of HPC i (1 to 5), in the month j (1 to 12) and year k (2003 or 2004), with the method l (CI, PV).

Detection probability of violations and of nonviolations is the probability that a future violation or lack of violation will be identified by the testing method. It was estimated by running 2 separate logistic regressions. Each model had the test results as the outcome. Detection probability of violations was estimated by modeling the probability of positive result (CIT+ or PVT+) using the subset of violations (V+; Dohoo et al., 2003). Detection probability of nonviolations was estimated by modeling the probability of negative result (CIT- or PVT-) using the subset of nonviolation (V-).

Certainty associated with a result is the probability that result of the testing method is correct. To estimate certainties associated with a positive and negative result 2 logistic regression analyses were run. Each had the true status (V+ or V-) as the outcome. To estimate certainty associated with a positive result, probability of violation (V+) was modeled and only positive results (CIT+ or PVT+) were included in the model. To estimate certainty associated with a negative result, probability of no violation (V-) was modeled and only negative results (CIT- or PVT-) were included in the model. Backward selection based on $P > 0.05$ criterion was used to eliminate insignificant terms from each model.

SPC Charts and CI as a Warning System

For each herd that exceeded a specific SCC level by at least 1 individual BTSCC, the date of the first violation was noted and the herd was classified as V+ for the SCC standard. Bulk tank SCC from the 30 d preceding that date were plotted on Shewhart's individuals

chart of SAS. Whenever any of the plotted BTSCC values met any of the rules outlined below, it was considered a signal of significant change in process output (BTSCC):

- A: A point more than 3 sigma away from the mean.
- B: Nine points in a row on one side of the mean.
- C: Six points in a row, all increasing or all decreasing.
- D: Two out of 3 points in a row 2 or more sigma away from the mean on one side of the mean.
- E: Four out of 5 points in a row 1 or more sigma away from the mean on one side of the mean.

Rules applied were chosen from run rules most commonly used in the manufacturing industry to detect changes in the process performance (Montgomery, 2005). To be included in this analysis, herds had to have at least 15 BTSCC values below the specific standard since the beginning of the monitoring period until they exceeded the standard. Analysis was done for each SCC standard separately. First the fraction of herds, for which the SPC charts showed a significant change in BTSCC during the 30 d preceding the SCC violation, was determined. This percentage represented the efficiency of detecting signals of an upcoming violation, when using the SPC charts alone (without the CI).

For the remaining herds (for which the SPC charts did not signal a change) the mean BTSCC during the 30 d preceding the standard violation was used to determine the CI [$CI = (\text{standard} - \text{mean})/3$]. To classify herds as CIT+ or CIT-, the calculated CI (maximum variation allowed) was then compared with the sigma observed during the same 30 d. The ability of the CI to correctly identify herds that will exceed a specific SCC level in the future was determined by finding the fraction of herds that were classified as CIT+, among those that exceeded the specific BTSCC level (V+), but for which the SPC chart did not signal. The fraction of herds, which would be warned of an upcoming BTSCC standard violation, within 30 d before it occurs, by the SPC chart signal or CI, was then calculated.

RESULTS AND DISCUSSION

CI Chart

For many herds, SPC charts along with a current BTSCC mean and sigma are being automatically computed from each bulk tank pickup test result (Kinsel, 2004) and provided as a service (via the Internet) to the dairy managers and dairy consultants. Knowledge of the mean along with the degree of variation (sigma) allows a quick assessment of milk quality process capability, and the dairy's ability to meet a specific standard.

Figure 2. Percentage of herds identified by the consistency index (CI) as capable of meeting a specific tank SCC (BTSCC) standard. Data presented for 5 SCC standards and 5 herd production categories (HPC). From left, bars represent HPC #1 (black bar), HPC #2 (white bar), HPC #3 (diagonally striped bar), HPC #4 (vertical striped bar), and HPC #5 (horizontal striped bar).

The CI developed in this study identifies herds able to consistently meet a specific SCC standard. The SCC standard adopted by the world's major exporters of milk and milk products (European Union countries and New Zealand) is 400,000 cells/mL. Based on the CI only 27.5% of the 1,501 herds were capable of consistently producing milk below the 400,000 cells/mL standard (Figure 2).

Using the CI chart (Figure 3) dairy managers and milk plant field staff can assess herds' capability of meeting one of the specified SCC levels. The first step involves determining herd's current BTSCC mean and sigma by use of the online service mentioned above or by calculating these estimates independently. Then the CI for any of the 5 SCC levels can be obtained from the CI chart. The CI value represents the maximum sigma, which can be allowed in BTSCC, to assure that chosen SCC level is consistently met. If the BTSCC sigma observed in the herd is above the CI value read off the CI chart, then the herd will not be able to consistently

Figure 3. Consistency index (CI) for the 200,000 cells/mL (■), 300,000 cells/mL (◆), 400,000 cells/mL (▲), 500,000 cells/mL (●), and 600,000 cells/mL (x) standard by tank SCC (BTSCC) mean.

produce milk below the desired SCC level. Therefore, based on the CI, an incapable herd can be identified even before a standard violation occurs. This makes the CI a proactive tool, encouraging process improvement before a standard violation causes loss of premium to the farmer, and significant decrease in milk processing value to the processor.

A worksheet was developed as an appendix to this manuscript. It calculates the CI based on BTSCC data provided by the user. The developed worksheet along with information about how to interpret its output is available online (<http://www.ansci.umn.edu/dairy/quality%20counts/SPREADSHEETS/MonitoringBTSCC.xls>).

SPC Charts and Indices as a Warning System

To be able to sell milk to the processor or receive a quality premium, the milk producer is expected to meet specific quality standards. Bulk tank SCC is one of those standards (NCIMS 2004). It is important, from a dairy manager's standpoint, to be able to determine whether the desired goal BTSCC can be assured, with the current level of the farm management. In this context, a CI along with SPC charts, could serve as a warning tool, predicting the possibility of crossing a specific SCC level, based on current BTSCC data. Routinely plotted BTSCC SPC process behavior charts can be used to monitor the compliance and consistency of all the processes that affect SCC level (Marsh et al., 1997; Reneau and Kinsel, 2001). Previous research has examined the use of SPC charts to detect changes in mastitis prevalence in the herd (Lukas et al., 2005) or changes to the milking routine (Reneau, 2000).

The index developed in this paper was derived from the capability index. Capability index is an SPC tool designed to evaluate the performance of a stable process (Montgomery, 2005). This means a process that does not experience changes to factors influencing its output. These factors might include the inputs, method, environment, operators, or machines. A chart that signals because of any of these special causes is therefore characteristic of an unstable process. To address this issue, our analysis of the CI was based on conditional probabilities. The values cited in Table 1 are probabilities of identifying a future violation given that herds' SPC charts did not signal during the 30 d preceding the violation.

Routine monitoring of SPC charts for signals of change would warn around 40% (Table 1) of the herds that a significant increase in BTSCC is occurring and may lead to a SCC standard violation. From the remaining herds (the herds that did not signal during the 30 d preceding a standard violation) another 44 to over

Table 1. Performance of the statistical process control (SPC) charts and consistency indices (CI) as a warning system

SCC standard cells/mL	n ¹	SPC chart, ² %	CI, ³ %	Either, ⁴ %
200,000	59	40.7	65.7	79.7
300,000	204	38.7	46.2	65.7
400,000	378	43.4	52.8	73.3
500,000	474	42.6	47.1	69.6
600,000	470	41.9	44.0	67.5

¹Number of herds included in the analysis.

²Percentage of herds for which the SPC chart signaled an increase in mean within 30 d before the SCC standard violation.

³Percentage of herds with their BTSCC in control (no signal on the SPC chart), for which the CI calculated for the 30 d before the SCC standard violation indicated a possibility of exceeding a specific SCC limit.

⁴Percentage of herds that would be warned either by the SPC charts or the CI of an upcoming SCC standard violation, within 30 d before a specific SCC limit is exceeded.

60% would be warned by the CI. For those herds the CI would correctly indicate that although their process is in control by SPC definition, a violation is likely to occur because the process is not capable. Lack of capability in this case would mean that for the mean BTSCC, which the farm is currently achieving, the process is not consistent enough to meet the desired SCC standard.

This results in 66 to 80% of the herds being warned by a control chart or the CI that a standard violation will occur before it actually occurs (Table 1). The best results (the highest percentage of herds warned before exceeding the SCC level) were obtained for the 200,000 cells/mL standard. These estimates, however, might not be as accurate and must be interpreted with caution because they are based on a smaller number of herds (n = 59). Statistical process control charts and CI are computed based on past performance data and can only reflect changes/causes that have already occurred (Montgomery, 2005). Therefore, by definition a CI or SPC chart will never be able to identify all herds that will exceed a specific SCC level just based on their past performance. Some herds are apt to experience sporadic changes in any of the factors affecting BTSCC. In small herds, for example, a single mastitis-infected cow, when undiagnosed and milked into the bulk tank, may cause a sudden BTSCC shift and an immediate SCC standard violation.

Percentage of Herds Correctly Identified

The PV identifies herds that are likely to violate the standard based on their past violations. This method is likely to be the current approach taken by milk processors wanting to ensure that their milk suppliers keep a specific SCC standard. Because BTSCC are subject

Figure 4. Detection probability of violations (□) and nonviolations (■) and certainty associated with a positive (△) and negative (▲) result of the consistency index (CI) test identifying herds violating the 400,000 cells/mL bulk tank SCC (BTSCC) standard.

to seasonal changes, analysis was performed on a monthly basis.

Percentage of herds correctly identified (as violators or nonviolators) did not differ between the 2 test methods, except for the 200,000 cells/mL standard, where it was 0.19% greater for the CI method (data not shown). The percent herds correctly identified is a function of the certainty associated with a positive and negative result and prevalence of standard violation. Greater prevalence of standard violations is associated with summer months and lesser SCC standard (Lukas et al., 2008). For both test methods (CI and PV), high certainty associated with a positive result is achieved at high prevalence, and high certainty associated with a negative result at low prevalence (Figures 4 and 5 and Tables

Figure 5. Detection probability of all nonviolations (DPV-) and violations (DPV+) and the certainty associated with a negative (CAT-) and positive result (CAT+) of the 400,000 cells/mL SCC standard. Results presented for the consistency index (CI) test. From left, bars represent HPC #1 (black bar), HPC #2 (white bar), HPC #3 (diagonally striped bar), HPC #4 (vertical striped bar), HPC #5 (horizontal striped bar). *Bars represent HPC for which probabilities are significantly different from HPC #5 ($P < 0.05$).

Table 2. Detection probability of violators (DPV+) and nonviolators (DPV-) in following month based on the consistency index (CI) or the past violations (PV) for the current month

Standard cells/mL	DPV+		DPV-	
	CI, %	PV, %	CI, %	PV, %
200,000	98.2 ^a	97.5 ^b	50.9 ^a	56.7 ^b
300,000	93.9 ^a	91.9 ^b	63.9 ^a	69.1 ^b
400,000	88.4 ^a	84.3 ^b	74.2 ^a	77.9 ^b
500,000	81.8 ^a	76.5 ^b	80.7 ^a	84.1 ^b
600,000	74.2 ^a	66.8 ^b	85.6 ^a	88.4 ^b

^{a,b}Values with different superscript within DPV category are different ($P < 0.05$).

2 and 3). This relationship explains the high percent correctly identified estimates (81 to 97%) for both methods.

Detection Probability and Certainty Associated with a Result

When evaluating and comparing diagnostic tests, it is important to distinguish between tests' ability to detect all violations as well as their ability to detect all nonviolations. Depending on the purpose of testing, it might be beneficial to favor one test over another regarding their ability to accurately detect BTSCC standard violators or nonviolators. Similarly, depending on the purpose of a diagnostics test, selection of a test with greater specificity may be favored over one with greater sensitivity (Martin et al., 1987; Dohoo et al., 2003).

One potential value of introducing the CI is its ability to identify herds that are incapable of meeting a specific SCC level by analyzing their past performance. This is of great value to milk processing plants procuring milk. The CI can help screen out those producers that will not be able to meet SCC quality standards. It can also prove helpful in segregating milk from individual farms before processing, based on farms' potential to produce milk below a specific SCC level.

Table 3. Certainty associated with a positive (CAT+) and negative (CAT-) test result when predicting violations in the following month based on the consistency index (CI) or the past violations (PV) for the current month

Standard cells/mL	CAT+		CAT-	
	CI, %	PV, %	CI, %	PV, %
200,000	97.1 ^a	97.4 ^b	62.5 ^a	57.4 ^b
300,000	90.4 ^a	91.5 ^b	73.5 ^a	69.2 ^b
400,000	82.2 ^a	84.0 ^b	81.8 ^a	78.0 ^b
500,000	73.4 ^a	76.2 ^b	86.8 ^a	84.2 ^b
600,000	64.6 ^a	68.1 ^b	90.4 ^a	88.3 ^b

^{a,b}Values with different superscript within CAT category are different ($P < 0.05$).

In a dairy farm setting, the cost associated with a false negative or failure to detect a violation may include a decrease in milk quality leading to loss of premium. Increased incidence of mastitis can also result from failure to detect signs of significant BTSCC increases. From a milk processor perspective, a false negative may cause loss because of decreased milk shelf life or processing value. The cost of a failure to detect nonviolations or a false positive on a farm will depend on the steps taken when a potential for future violation is identified. If a positive result of a CI or PV test will simply lead to an increased vigilance in milking or bedding routine, then the cost will be limited to the cost of additional material or labor and may actually lead to improved human or animal performance. As the number of false positives increase, however, there is an increasing risk that the results of a test will be ignored. Therefore, for a test to serve its purpose, it has to be characterized by a high detection probability of violations and an acceptable certainty associated with a positive result.

Prevalence of standard violations was not predetermined by study design, varied for different standards and HPC, and changed throughout the study period (Lukas et al., 2008). This gives an opportunity to calculate and compare the certainty associated with a result at different prevalence levels for each of the test methods (Dohoo et al., 2003).

The full logistic regression models, used to estimate certainties associated with a result and detection probabilities, included 4 main factors (method, HPC, month, and year) and all the 2-way interactions with method (HPC $\times$ method, month $\times$ method, and year $\times$ method). The only 4 significant interactions were the method $\times$ HPC interaction in the detection probability of all nonviolations of the 400,000, 500,000, and 600,000 cells/mL standard model; and the method $\times$ year interaction for the detection probability of all nonviolations of the 600,000 cells/mL standard model (data not shown). The inclusion of the significant interaction terms in the estimation of detection probability of nonviolators would change the estimate by a maximum of 1.3%. Considering the limited impact of the interaction on the comparisons of the 2 test methods, and to simplify the analysis, the interactions were dropped. Only main factors were considered in further interpretation of the results.

Consistency index detection probabilities and certainties associated with a result for the 400,000 cells/mL standard are presented by month (Figure 4) and by HPC (Figure 5). Similar patterns were observed for the remaining 4 standards and the PV method (data not shown).

The differences between methods and HPC were significant for all 4 measures and across all 5 BTSCC levels (Tables 2 and 3). Detection probability of all violators and certainty associated with a positive result decreased as standard increased. Inversely, detection probability of all nonviolators and certainty associated with a negative result increased with standard. Detection probability of all violators, across all BTSCC standards, was from 0.7 to 7.4% greater for the CI (Table 2). Similarly, certainty associated with a negative result was from 2.1 to 5.1% greater for the CI than for the PV method (Table 3). The opposite was true for detection probability of all nonviolations (2.7 to 5.8% greater for the PV method, Table 2) and certainty associated with a positive result (0.3 to 3.5% greater for the PV method, Table 3).

Detection probability of nonviolations and certainty associated with a negative result was greater in 2004 than 2003 across all BTSCC levels. Detection probability of violations and certainty associated with a positive result was less in 2004 than 2003 for most standards (data not shown). The detection probability of violations of 500,000 cells/mL standard and nonviolations of 200,000 cells/mL standard did not change significantly with year. The year factor was also removed from models estimating certainty associated with a positive result for 500,000 cells/mL standard and negative result for 200,000 and 600,000 cells/mL standard. Including the year factor in the models changed the estimates by a maximum of 3.5% for both methods equally because apart from the 600,000 cells/mL standard, the year $\times$ method interaction was not significant. The relationship between year and certainty associated with a result is a consequence of greater prevalence of violations in 2003 (Lukas et al., 2008).

All of the test performance measures for all standards change ($P < 0.05$) with month. The detection probability of nonviolations and certainty associated with a positive result decreased from July to October (Figure 4). Season has the opposite relationship with detection probability of violations and certainty associated with a negative result (Figure 4). This pattern was repeated across all standards (data not shown). The cyclic pattern in the detection probability of all violations and certainty associated with a positive result was even more apparent in the greater BTSCC standards (500,000 and 600,000 cells/mL). For the lesser BTSCC standards (200,000 and 300,000 cells/mL), detection probability of nonviolations and certainty associated with a negative result was more visibly impacted by month (data not shown).

A significant increase in the number of violations (i.e., prevalence) is observed in the summer months and returns to lesser values in fall for all of the 5 BTSCC standards (Lukas et al., 2008). Therefore, it can be ex-

pected that a fraction of herds identified as positive (CIT+ or PVT+) by either of the 2 methods in summer will improve (by decreasing their mean or variation in BTSCC) and not exceed the limit in fall months. This will have a negative impact on the detection of all nonviolations and certainty associated with a positive result of both test methods when performed in the late summer or early fall months (August, September, and October). It will also improve the detection of all violations and certainty associated with a negative result, during that same period. Prevalence is less, but changes in prevalence associated with season are greater for the greater BTSCC levels (Lukas et al., 2008). Therefore, the cyclic pattern in detection of all violations and certainty associated with a negative result (Figure 4) will be more apparent for the greater standards.

The greater the BTSCC standard, the easier it is for the herd to stop violations because of changes in management or as a result of the effect of season on the BTSCC. This results in an increase in the number of false positive with increase in SCC standard, causing certainty associated with a positive result to drop (Table 3). The opposite is true for certainty associated with a negative result. The lesser the standard, the harder it is for a herd to consistently keep it. This increases the number of false negatives as the BTSCC standard drops, causing a drop in certainty associated with a negative result (Table 3). This relationship mimics and is magnified by the effect of prevalence on certainty associated with a result (Martin et al., 1987).

The lesser the standard the herd attempts to meet, the more intense management it will require. The more intense the management, the more likely for the herd manager to decide to use a CI or PV method to assess prospects of meeting a desired BTSCC standard. For the greater BTSCC standards, 500,000 and 600,000 cells/mL, the certainty associated with a positive result drops to a low of 64.6%. For the lesser standards, however, it is maintained above 82%. This makes a positive result of either of the tests (CI or PV) reliable on farms where the goal is to keep BTSCC always below 400,000 cells/mL.

Detection probability of violations and certainty associated with a positive result decrease with the increase in HPC (Figure 5). The opposite is true for detection probability of nonviolations and certainty associated with a negative result. The magnitude of change with HPC, however, increased with standard (data not shown). Certainty associated with a result is known to be driven by prevalence (Martin et al., 1987), which in our data set was also impacted more by HPC and season at greater standards (Lukas et al., 2008). The change in detection probability with HPC can be explained by the differences in BTSCC mean and sigma between

HPC. The larger the herd, the less variation is observed in BTSCC (Lukas et al., 2005, 2008). This also means that larger herds can maintain their BTSCC closer to a standard without crossing it. Small herds, which naturally have large variation, are likely to repeatedly violate a BTSCC standard each month. This causes both methods to be more lenient for larger herds that naturally will tend to have low variation. This might partially explain the decrease in the detection probability of all violations as the HPC increases.

Detection probability of a diagnostic test can be manipulated by selecting the critical SCC threshold (Martin et al., 1987). In this study the threshold level is the SCC standard the herd is expected to meet the following month. Adjusting the threshold to account for the relationship with season may, however, prove beneficial. A significant increase in mean and variation naturally occurs in most of the herds in the summer months (Lukas et al., 2008). Therefore, the decrease in detection probability of all nonviolations in September can be limited by selecting a greater threshold for that month. Similarly the drop of detection probability of all nonviolations in the summer can be reduced by selecting a lesser threshold level for the late spring and early summer months. This means that a herd classified as CIT- or PVT- for June would have to have maintained lesser BTSCC (mean or variation) in the previous month compared with a herd classified CIT- or PV- for September. The relationship between certainty associated with a result, and detection probability and HPC can be accounted for in the same manner. Because of the identified difference in test performance between years and possible interactions between month and HPC, the exact selection of threshold should be the subject of a follow-up study when data from more than 2 yr are available.

The greater performance of the CI, in terms of detection probability of all violators and certainty associated with a negative result, indicates that the CI method is better in identifying all the violating herds. It also indicates that a negative result of the CI test (classifying a herd as a nonviolation) is more reliable than screening herds based on their past violations.

For a PV method to detect a future violator, the herd has to first violate a given standard. The fact that a true future standard violator can be identified by a CI method even if a herd's BTSCC never crossed the desired SCC level makes this method a more proactive management tool.

The analysis presented in this study gives an opportunity to associate a positive or negative result with a level of certainty depending on the size of the herd production and month of testing. This additional information makes decision making on the farm more fact

based and flexible to the management style and the level of risk the manager decides to take. Knowledge of the level of certainty associated with a result will influence the decision on the type of action initiated in response to a positive or negative result. Identification of other factors that influence prevalence could further improve the estimate of level of certainty and could be the subject of further study.

CONCLUSIONS

The SPC tools presented in this study help the farmer decide whether corrective action is needed to prevent future standard violations. They can also help assess whether preventive action was successful in decreasing the mean or variation in BTSCC. By outperforming the PV method, the CI can also better serve the milk processor, when identifying herds that are more likely to consistently keep a specific SCC standard.

The goal of maintaining the BTSCC of domestic raw milk supply below 400,000 cells/mL must be sought if the US dairy industry desires to be a serious competitor in global dairy markets. In this research, based on the CI, only 27.5% of the 1,501 herds could successfully meet this standard. The CI in combination with the SPC chart plotted for the BTSCC offer a proactive approach to maintaining consistently high milk quality. They allow assessing process capability and distinguishing between significant changes and random variation in BTSCC.

REFERENCES

- Barkema, H. W., Y. H. Schukken, T. J. Lam, M. L. Beiboer, G. Benedictus, and A. Brand. 1998. Management practices associated with low, medium, and high somatic cell counts in bulk milk. *J. Dairy Sci.* 81:1917–1927.
- Chassagne, M., J. Barnouin, and M. Le Guenic. 2005. Expert assessment study of milking and hygiene practices characterizing very low somatic cell score herds in France. *J. Dairy Sci.* 88:1909–1916.
- de Vries, A., and B. J. Conlin. 2003. Economic value of timely determination of unexpected decreases in detection of estrus using control charts. *J. Dairy Sci.* 86:3516–3526.
- Dohoo, I., S. W. Martin, and H. Stryhn. 2003. Screening and diagnostic tests. Pages 85–121 in *Veterinary Epidemiology Research*. AVC Inc., Prince Edward Island, Canada.
- Eberhart, R. J., L. J. Hutchinson, and S. B. Spencer. 1982. Relationship of bulk tank somatic cell counts to prevalence of intramammary infection and indices of herd production. *J. Food Prot.* 45:1125–1128.
- Kinsel, M. L. 2004. Understanding variation: The key to reducing bulk tank somatic cell count. *Proc. Natl. Mastitis Council. Regional Meet.*, Bloomington, MN.
- Kniffen, T. 1994. Potential uses of SPC in a pork production system. Pages 1–12 in *Statistical Process Control*. *Proc. Allen D. Leman Swine Conf.* University of Minnesota, St. Paul, MN.
- Lukas, J. M., D. M. Hawkins, M. L. Kinsel, and J. K. Reneau. 2005. Bulk tank somatic cell counts analyzed by statistical process control. *J. Dairy Sci.* 88:3944–3952.
- Lukas, J. M., M. L. Kinsel, and J. K. Reneau. 2008. Predicting SCC standard violations based on herd's bulk tank SCC. Part I: Analyzing variation. *J. Dairy Sci.* 91:427–432.
- Marsh, W. E., A. de Vries, J. K. Reneau, and M. L. Kinsel. 1997. Monitoring performance: statistical process control in dairy herd management. Pages 34–46 in *Annual Northeast Dairy Production Medicine Symposium*. 6th ed. Syracuse, NY.
- Martin, S. W., A. H. Meek, and P. Willenberg. 1987. Measurements of disease frequency and production. Pages 48–76 in *Veterinary Epidemiology Principles and Methods*. Iowa State Univ. Press, Ames, IA.
- Montgomery, D. C. 2005. *Introduction to Statistical Quality Control*. 5th ed. J. Welter, J. Powers, S. Dumas, D. Stanley, and T. Kulesa. John Wiley & Sons Inc., Hoboken, NJ.
- National Animal Health Monitoring System. 2003. Dairy 2002 Part II: Changes in the United States Dairy Industry, 1991–2002. <http://www.aphis.usda.gov/vs/ceah/ncahs/nahms/dairy> Accessed Aug. 13, 2007.
- National Conference on Interstate Milk Shipments. 2004. Grade “A” Pasteurized Milk Ordinance 2003 Revision. Public Health Service/Food and Drug Admin. Online. <http://vm.cfsan.fda.gov/~ear/p-nci.html> Accessed Aug. 13, 2007.
- Niza-Ribeiro, J., J. P. T. M. Noordhuizen, and J. C. Menezes. 2004. Capability index—A statistical process control tool to aid in udder health control in dairy herds. *J. Dairy Sci.* 87:2459–2467.
- Reneau, J. K. 2000. Process control: Timely feedback for quality milk production at the farm. *Proc. National Mastitis Council Annual Meet.*, Atlanta, GA. Natl. Mastitis Council. Inc., Madison, WI.
- Reneau, J. K., and M. L. Kinsel. 2001. Record systems and herd monitoring in production-oriented health and management programs in food producing animals. Pages 107–146 in *Herd Health*. O. M. Radostits, ed. W. B. Saunders Company, Philadelphia, PA.
- Reneau, J. K., A. J. Seykora, B. J. Heins, M. I. Endres, R. J. Farnsworth, and R. F. Bey. 2005. Association between hygiene scores and somatic cell scores in dairy cattle. *J. Am. Vet. Med. Assoc.* 227:1297–1301.
- Sard, D. M. 1979. Dealing with data: The practical use of numerical information—(14) Monitoring changes. *Vet. Rec.* 105:323–328.