Research Article

Quality and Reliability Engineering International

(wileyonlinelibrary.com) DOI: 10.1002/qre.1971

Published online in Wiley Online Library

The *c*-Chart with Bootstrap Adjusted Control Limits to Improve Conditional Performance

Meng J. Zhao*† and Anne R. Driscoll

The integrity of Phase II control charting depends on the accuracy of Phase I estimation. Studies have shown that extremely large sample sizes are needed in Phase I to ensure that performance of control charts with estimated in-control parameters is comparable with the performance of charts with known parameters. The sample size recommendations can be impractical for attribute control charts. In this article, the in-control performance of the *c*-chart with an estimated in-control average number of non-conforming items is assessed. We show that the sampling variability associated with estimation results in a high percentage of control charts with in-control average run lengths well below that of corresponding control charts with known parameters. This sampling variability can be described as between-practitioner variability. To overcome the variability in performance, a *c*-chart with bootstrapped control limits is recommended. A simulation study reveals that these adjusted bootstrapped control limits improve the conditional average run length performance of the *c*-chart by controlling the proportion of charts with in-control average run length performance below a given value. The out-of-control performance of the *c*-chart with adjusted limits is also discussed. Copyright © 2016 John Wiley & Sons, Ltd.

Keywords: attribute control charts; AARL; estimation effect; SDARL; standard deviation of average run length; statistical process control

1. Introduction

When choosing the appropriate control chart for monitoring, one must consider the type of data being collected. Attribute data, such as counts or pass/fail data, are often collected and much research has been conducted for determining the appropriate monitoring methods for these types of data. Woodall[1] and Topalidou and Psarakis[2] provided reviews of the literature on monitoring techniques for attribute type data.

In quality control, the *c*-chart is the standard attribute control chart used to monitor the number of defects in a manufacturing process. The number of non-conformities in a given sampling period or in a given defect prone part of the process is plotted on a *c*-chart to monitor process stability. In order to determine the appropriate control limits for the *c*-chart, the expected number of non-conformities for the in-control process must be known. However, this parameter is often unknown, and in order to monitor a process in real time, the parameter must first be estimated. This estimation occurs in Phase I of the control charting process. Phase I is a retrospective analysis of process data where these parameters can be estimated to establish the control limits to be used in future monitoring of the process. Additional objectives in Phase I include understanding the process behavior and determining if the process is in statistical control. Phase II follows Phase I, and it is in this phase where real-time monitoring of the process occurs with the objective of quick detection in out-of-control conditions. Chakraborti *et al.*[3] and Jones-Farmer *et al.*[4] provided overviews of Phase I methods and objectives.

The integrity of Phase II monitoring depends on the reliability of Phase I estimation. The performance of control charts with estimated in-control process parameters can be much different than for charts where these process parameters are known. Research has been conducted on the effects of estimation on attribute control charting. Braun[5] studied run length properties for attribute charts with estimated parameters. Chakraborti and Human[6,7] evaluated the effect of parameter estimation on the performance of the *p*-chart and the *c*-chart. Raubenheimer and Merwe[8] also extended the work of Chakraborti and Human[7] by incorporating a Bayesian approach to study *c*-charts for monitoring non-conformities and comparing the Bayesian charts to the traditional *c*-charts with three-sigma limits studied by Chakraborti and Human[7]. Raubenheimer and Merwe[8] determined that the Jeffery's prior produced the best convergence rate of the posterior predictive densities of the process parameter. They studied various parameter values and sample sizes. Chakraborti and Human[7] and Raubenheimer and Merwe[8] constructed the *c*-charts with three-sigma limits. Their results indicated that the *c*-chart performance is poor when the true average number of non-conformities is small, which is expected. When

Department of Statistics, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA
**Correspondence to: Meng John Zhao, Virginia Polytechnic Institute and State University, Department of Statistics, Blacksburg, VA, USA.*
†E-mail: jmz167@vt.edu

using three-sigma limits in situations with a small parameter value, the lower control limit often holds no meaning because it is negative, that is, $c < 3\times\sqrt{c}$. The distribution is also highly skewed in these instances, which further invalidates the use of symmetric three-sigma limits. In our research, we determine that probability limits are more appropriate for *c*-charts that monitor Poisson random variables. We will use probability limits when determining the control limits for the *c*-chart to overcome these problems.

Castagliola and Wu[9] extended the work on estimation error for attribute control charting by evaluating the necessary sample sizes for the *c*-chart and *p*-chart in Phase I analysis where the upper and lower control limits are based on the three-sigma limits approach. The authors also provided tables with parameters to adjust the three-sigma limits to improve chart performance. These parameters depend on the Phase I sample size. Castagliola *et al.*[10] studied the effect of synthetic attribute charts with estimated parameters. Testik[11] explored the effect of estimation error on the Poisson cumulative sum chart and provided sample size recommendations. Jensen *et al.*[12] and Psarakis *et al.*[13] provided reviews of the literature on performance of charts with estimated parameters. For many of these methods, the sample size recommendations to overcome the effect of estimation error are impractical. The overall conclusion from the research is that estimation of parameters can have a strong adverse effect on chart performance.

A common metric for evaluating the performance of control charts during Phase II is the average run length (ARL). The ARL of a control chart is the average number of statistics that are plotted before a chart signals. This signal can correctly indicate a change in the process or be a false alarm, meaning the chart signaled when, in fact, there was no shift. The out-of-control ARL is the expected number of samples observed before a shift is detected, while the in-control ARL is the expected number of samples observed before a false alarm. Typically, the goal is to minimize the out-of-control ARL over a range of process shifts while requiring a specified value for the in-control ARL.

The process of designing a *c*-chart is similar to other types of control charts. In practice, the expected number of non-conformities in the process is estimated from a Phase I sample. When this estimation occurs, the Phase II control limits become random variables. Because the control limits are determined based on the estimated parameter and a chosen false alarm rate, the performance of the charts can degrade significantly depending on the value of the parameter estimates (Chakraborti and Human[7]). It is well known that the performance of control charts can vary depending on the accuracy of the estimated parameters.

Castagliola and Wu[9] studied the unconditional run length properties of *c*-charts with estimated parameters and compared the in-control ARL performance of these charts with charts where the parameters are known. The *c*-charts were constructed with three-sigma upper and lower control limits. Castagliola and Wu[9] recommended sample sizes necessary to ensure the relative difference between the in-control ARL for charts with estimated parameters and with known parameters is no larger than 5%. In some situations, the necessary Phase I samples can be larger than 10,000. The discreteness of the data results in sample sizes that change drastically depending on the in-control parameters for the *c*-chart. There was also no clear pattern in terms of the recommended sample sizes. In many cases, the Phase I sample size recommendation are impractical.

Recently, researchers have focused on designing charts to overcome the challenges associated with estimation error. The evaluation of control charting methods has shifted from the unconditional ARL metric to the conditional ARL metric. The unconditional ARL metric averages over the sampling variability that results from Phase I estimation. The average of the unconditional ARL metric is referred to as average ARL. A better metric that illustrates the amount of sampling variability is the standard deviation of the ARL. Saleh *et al.*[14] referred to this sampling variability as practitioner-to-practitioner variability. They commented that estimates of process parameters will differ among practitioners because each practitioner will be drawing his or her own Phase I sample from the underlying distribution of the process. Practitioners should be concerned with properties of their particular control chart, not necessarily with the unconditional performance that is averaged across practitioner-to-practitioner variability. Saleh *et al.*[14] studied the standard deviation of the ARL of the $\overline{X}$ control charts and X charts for individuals and concluded that far larger samples sizes than previously recommended are necessary to reduce the effect of estimation error when the practitioner-to-practitioner variability is taken into account. Albers and Kallenberg[15] were the first researchers to recognize the issue of sampling variability.

Alternative techniques to account for practitioner-to-practitioner variability have been proposed in the literature. Designing control charts with control limits that result from bootstrap simulations have been suggested to be a viable way to eliminate the need for unpractically large Phase I samples and also reduce the adverse impact of practitioner-to-practitioner variability. The parametric bootstrap approach was proposed by Jones and Steiner[16] and Gandy and Kvaløy[17]. To design charts that achieve a certain degree of conditional performance, they adjusted the control limits to ensure that a specified percentage of in-control ARL values were at a certain level or higher. A bootstrap approach was used to determine the adjusted control limits. Saleh *et al.*[18] studied the conditional performance of the exponentially weighted moving average control chart and used the bootstrap approach to design exponentially weighted moving average charts with adjusted control limits.

The majority of existing studies on dealing with estimation errors for *c*-chart have three-sigma limits. However, when constructing *c*-charts, there is a difference in performance for the chart when using three-sigma limits versus probability limits. Especially in cases where the c_0 value is small, the use of symmetric limits does not make sense. For this reason, we focus our studies on charts with probability limits. In this article, we first evaluate the conditional performance of the *c*-chart with probability limits to understand the effect of practitioner-to-practitioner variability. We then extend the parametric bootstrap method to adjust the control limits of the *c*-chart and study the out-of-control performance. We show that the bootstrap method improves the in-control performance of the *c*-chart.

In Section 2, the construction of *c*-chart with known in-control parameters is discussed, followed by an overview of the *c*-chart with estimated in-control parameters in Section 3. The conditional in-control ARL performance for the *c*-chart is presented in Section 4. In Section 5, we introduce the parametric bootstrap method for obtaining adjusted control limits for *c*-charts with estimated parameters. We then present and compare the ARL performance of our adjusted method with the ARL performance resulting from the use of unadjusted control limits in Section 6. In addition, we also consider the out-of-control performance when such adjustments are made in Section 7. The last section, Section 8, provides our concluding remarks.

2. The *c*-chart with known in-control parameters

A common approach to modeling count data is to use the Poisson distribution. Let $X_1, X_2, \ldots$ be a sequence of observed counts of the number of non-conformities. We assume that $X_1, X_2, \ldots$ are independent Poisson observations and that X_i has an in-control mean of c_0.

The Shewhart *c*-chart is the most common control chart for monitoring Poisson distributed data. The number of non-conformities X_i is plotted on the *c*-chart and the chart signals when the observed count reaches or exceeds the lower or upper control limits. The upper control limits (UCL_{c_0}) and lower control limits (LCL_{c_0}) are found such that

$$\alpha \geq P(X \geq UCL_{c_0} \mid c_0) + P(X \leq LCL_{c_0} \mid c_0), \tag{1}$$

where α is the chosen false alarm rate of the chart and X is a Poisson random variable with parameter c_0, the in-control expected number of non-conforming items. The limits in Equation (1) are based on probability limits rather than the popular three-sigma limits. We chose to use the probability limits because for small values of c_0, the distribution is not symmetric.

Because X is a Poisson random variable, a consequence of the discreteness of X is that the exact desired value for false alarm rate can almost never be obtained. Akin to the calculation for confidence intervals, infinite combinations of UCL_{c_0} and LCL_{c_0} exist that satisfy Equation (1). In order to avoid confusion and to define a set of control limits to implement, we employ the following calculation for both control limits:

$$\alpha_{c_{0U}} = P(X \geq UCL_{c_0} | c_0) \leq \frac{\alpha}{2} \quad \textit{and} \quad \alpha_{c_{0L}} = P(X \leq LCL_{c_0} | c_0) \leq \frac{\alpha}{2}, \tag{2}$$

where $\alpha_{c_{0U}}$ is the false alarm rate for exceeding the upper control limit and $\alpha_{c_{0L}}$ is the false alarm rate for signals below the lower control limit. The resulting false alarm rate for the control chart, α_{c_0}, is

$$\alpha_{c_0} = \alpha_{c_{0U}} + \alpha_{c_{0L}}. \tag{3}$$

The false alarm rate, α_{C_0}, for the chart is probability that the control chart will signal when in fact the process is stable. This results in an observed in-control ARL value, $ARL_{c_0} = \frac{1}{\alpha_{c_0}}$.

When the value for c_0 is small, it is possible that no LCL_{c_0} value exists because $P(X=0) > \alpha$. If this is the case, then the value of UCL_{c_0} is found such that

$$\alpha_{c_{0U}} = P(X \geq UCL_{c_0} | c_0) \leq \alpha. \tag{4}$$

Table I illustrates our method for calculating the control limits for cases when $c_0 = 3$, 10, 20,and 50, and the desired value for the false alarm rate $\alpha = 0.01$. Equation (2) was used to calculate the control limits for the cases when $c_0 = 10$, 20, and 50. Because the lower control limit does not exist when $c_0 = 3$, Equation (4) was used to calculate only the upper control limit. The fourth and fifth columns in Table I give the corresponding lower and upper false alarm rates, and the sixth column is the overall false alarm rate. The desired ARL value is 100; however, all the cases result in ARL_{c_0} values that differ from 100 because of the discreteness of the data.

3. The *c*-chart with estimated in-control parameters

In practice, the value of c_0 is typically unknown and, as a result, the value of c_0 must be estimated from readily available data. Let $X_1, X_2, \ldots, X_m$ be a set of Phase I samples. Assuming that the process is stable, the estimate of c_0 is $\hat{c} = \frac{1}{m}\sum_{i=1}^{m} X_i$ for $i = 1, 2, \ldots, m$. With this estimator, the false alarm rate expressed in Equation (1) is adapted to incorporate $\hat{c}$ resulting in

$$\alpha \geq P(X \geq UCL_{\hat{c}} | \hat{c}) + P(X \leq LCL_{\hat{c}} | \hat{c}). \tag{5}$$

The method developed in Equation (2) for determining the control limits is adjusted for $\hat{c}$ such that

$$\alpha_{\hat{c}_U} = P(X \geq UCL_{\hat{c}} | \, \hat{c}) \leq \frac{\alpha}{2} \quad \text{and} \quad \alpha_{\hat{c}_L} = P(X \leq LCL_{\hat{c}} | \, \hat{c}) \leq \frac{\alpha}{2}, \tag{6}$$

Table I. Control limits, false alarm rates, and ARL_{c_0} values for *c*-charts with probability limits. The false alarm rate $\alpha = 0.01$ is desired

c_0	LCL_{c_0}	UCL_{c_0}	$\alpha_{c_{0L}}$	$\alpha_{c_{0U}}$	α_{c_0}	ARL_{c_0}
3	—	8	—	0.0038030	0.0038030	262.95
10	2	19	0.0027694	0.0034540	0.0062234	160.68
20	9	32	0.0049954	0.0047270	0.0097224	102.86
50	32	69	0.0043929	0.0043350	0.0087279	114.58

where $\alpha_{\hat{c}_U}$ and $\alpha_{\hat{c}_L}$ are the observed upper and lower false alarm rates for the charts with control limits based on $\hat{c}$. The overall false alarm rate, $\alpha_{\hat{c}}$, is

$$\alpha_{\hat{c}} = \alpha_{\hat{c}_U} + \alpha_{\hat{c}_L}. \tag{7}$$

For cases when the value for $\hat{c}$ is small and no lower control limit exists, the upper control limit will be

$$\alpha_{\hat{c}} = \alpha_{\hat{c}_U} = P(X \geq UCL_{\hat{c}}|\hat{c}) \leq \alpha, \tag{8}$$

where $\alpha_{\hat{c}U}$ is the upper false alarm rate, but in this case, this is equivalent to the overall false alarm rate $\alpha_{\hat{c}}$.

Up to this point, we have discussed the following three overall false alarm rates:

- α: the desired false alarm rate.
- α_{c_0}: the observed false alarm rate when c_0 is known.
- $\alpha_{\hat{c}}$: the observed false alarm rate when c_0 is estimated with $\hat{c}$.

There is a fourth false alarm rate that is necessary to define in order to explore the effect of estimation error on the performance of the c-chart. The control chart limits are defined for the case when $\hat{c}$ estimates c_0. These limits can be calculated using Equation (6). We then define $\alpha_{\hat{c}|c_0}$ as the false alarm rate for a control chart with control limits based on $\hat{c}$, but the probabilities are computed using the true mean value c_0. The false alarm rate is

$$\alpha_{\hat{c}|c_0} = P(X \geq UCL_{\hat{c}}|c_0) + P(X \leq LCL_{\hat{c}}|c_0). \tag{9}$$

This metric indicates the actual estimation error. The practitioner will never be aware of this value, but in simulation studies, we are able to calculate this error rate because we know the true value for c_0. We can then use the false alarm rate to calculate $ARL_{\hat{c}|c_0}$, which is the in-control ARL value for c-charts with limits computed using $\hat{c}$, but the overall c_0 value is known.

4. Conditional performance of the *c*-chart with estimated parameters

When the parameter c_0 is estimated, the in-control ARL becomes a random variable because of random Phase I sampling. Different Phase I samples from the same process will likely result in different upper and lower control limits because of sampling variability. In order to investigate the effect of parameter estimation on chart performance, a simulation study was performed to explore variability associated with in-control ARL performance. In the study, 10,000 Phase I samples were simulated with $m = 20$, 50, 100, 1000,and 5000 for $c_0 = 3$, 10, 20, and 50. The desired false alarm rate was chosen to be $\alpha = 0.01$. For each Phase I sample, the $LCL_{\hat{c}}$ values and $UCL_{\hat{c}}$ values were computed using Equation (6) except for the case when $c_0 = 3$. There is no lower control limit for this case so Equation (8) was used. Following the computation of the control limits, the $ARL_{\hat{c}|c_0}$ values were computed based on the in-control parameter c_0. Tables II and III show the in-control average $ARL_{\hat{c}|c_0}$ $\left(AARL_{\hat{c}|c_0}\right)$ and the in-control standard deviation of $ARL_{\hat{c}|c_0}$ $\left(SDARL_{\hat{c}|c_0}\right)$ for the various cases of the simulation.

Table II. In-control average $AARL_{\hat{c}|c_0}$ performance for varying Phase I sample sizes, m

	m					
c_0	20	50	100	1000	5000	∞
3	246.55	211.42	211.35	255.47	262.95	262.95
10	153.05	161.91	162.35	160.68	160.68	160.68
20	121.94	131.01	134.14	132.17	127.30	102.86
50	107.03	115.59	119.09	117.73	114.61	114.58

Table III. In-control $SDARL_{\hat{c}}|c_0$ values for varying Phase I sample sizes, m

	m					
c_0	20	50	100	1000	5000	∞
3	269.52	131.94	88.57	35.82	0.00	0.00
10	50.10	41.66	32.00	0.00	0.00	
20	27.06	19.54	15.48	17.32	20.62	
50	21.15	15.22	13.09	8.30	0.94	

As the Phase I sample size increases, the $AARL_{\hat{c}|c_0}$ values approach ARL_{c_0}, the in-control ARL value when c_0 is known. The ARL_{c_0} corresponds to the seventh column of Table II where the sample size $m = \infty$. For all cases of c_0 except $c_0 = 20$, the $AARL_{\hat{c}|c_0} = ARL_{c_0}$ when Phase I sample size $m = 5000$. For $c_0 = 20$, the $AARL_{\hat{c}|c_0}$ values are all larger than the in-control ARL value, ARL_0. A similar phenomenon was observed by Castagliola and Wu[9] in their study of the in-control ARL performance of the *c*-chart with estimated parameters.

As the Phase I sample size, *m*, increases, the $SDARL_{\hat{c}|c_0}$ values tend to decrease. However, we notice that the $SDARL_{\hat{c}|c_0}$ metric is not decreasing in a monotonic pattern as expected for the $\overline{X}$ chart. There are even cases where the $SDARL_{\hat{c}|c_0}$ values increase as *m* increases. For example, when $c_0 = 20$, the $SDARL_{\hat{c}|c_0}$ increases from 17.32 to 20.62 when *m* changes from 1000 to 5000. The reason the $SDARL_{\hat{c}|c_0}$ is still a rather large value when $m = 5000$ can be attributed to the discreteness of the data and the closeness of the control limits to $\alpha/2$. In contrast to three-sigma limits, when using probability limits, there exists an inherent robustness to estimation error because the estimated value $\hat{c}$ can vary a certain amount away from c_0 without causing any changes in the probability limits or in-control ARL performance. In the case of $c_0 = 20$, Table I quotes the $ARL_{c_0} = 102.86$. This values of c_0 results in a ARL_{c_0} closest to the desired $ARL = 100$ when $\alpha = 0.01$. This means that $\alpha_{c_{0L}}$ and $\alpha_{c_{0U}}$ must be very close to the desired $\frac{\alpha}{2} = 0.005$. Table I states for $c_0 = 20$, $\alpha_{c_{0L}} = 0.0049954$ and $\alpha_{c_{0U}} = 0.004727$. Therefore, small shifts in *c* from c_0 will move the upper or lower control limits to ensure the false alarm rates are below 0.005 resulting in a change in $ARL_{\hat{c}|c_0}$.

The $SDARL_{\hat{c}|c_0}$ metric illustrates the practitioner-to-practitioner variability. However, it is also of interest to determine the percentage of control charts with $ARL_{\hat{c}|c_0}$ below ARL_{c_0}. These control charts have a higher false alarm rate, $\alpha_{\hat{c}|c_0}$, compared with the desired value α_{c_0}. Using the same simulation study discussed previously with 10,000 Phase I samples for $m = 20$, 50, 100, 1000, and 5000 and $c_0 = 3$, 10, 20, and 50, the percentage of charts with $ARL_{\hat{c}|c_0}$ below ARL_{c_0} was recorded. The results of this simulation study are given in Table IV. For the different scenarios, the percentage of charts with $ARL_{\hat{c}|c_0}$ below the expected performance can be as high as 64.30% as illustrated when $c_0 = 50$ and $m = 20$. The percentage of underperforming charts decreases as the Phase I sample size increases, as expected. These ARL values reveal the effects of practitioner-to-practitioner variability. The practitioner computes the control limits based on specified in-control performance, but because of estimation error, the observed in-control ARL performance can deviate from the desired performance.

The variability in the conditional ARL performance, $ARL_{\hat{c}|c_0}$, can be attributed to the estimation error as a result of estimating c_0 with $\hat{c}$. From Table IV, it is shown that when the Phase I sample size $m = 5000$, the conditional ARL performance for all of the control charts reaches the desired ARL value, resulting in 0% of the charts with $ARL_{\hat{c}|c_0}$ values below the targeted ARL_{c_0} value.

In Figure 1, the conditional ARL distribution is displayed using histograms when $c_0 = 20$ for varying Phase I sample sizes. The targeted in-control ARL value for the *c*-chart with $c_0 = 20$ is 102.86, which is represented in the plots with a red reference line. These histograms illustrate that there is still variability in conditional ARL performance as a result of sampling variability. Only certain ARL values can be achieved with the chart because of the discreteness of the data. In Figure 1(d) when $m = 1000$ it is apparent that there is still variability in conditional ARL performance, but all conditional ARL values are above the targeted $ARL_0 = 102.86$.

In the next section, a parametric bootstrap method is described to decrease the percentage of charts with in-control conditional ARL values below the target ARL_{c_0} values. These bootstrapped control limits are recommended to help lessen the negative impact of practitioner-to-practitioner variability.

5. Bootstrap adjusted control limits for the *c*-chart

Previous studies have evaluated the effect of Phase I sampling variability has on the $AARL_{\hat{c}|c_0}$ value using three-sigma limits. However, the use of $AARL_{\hat{c}|c_0}$ as an evaluation criterion is not complete rate because the $AARL_{\hat{c}|c_0}$ value is a misleading metric when evaluating the performance of a *c*-chart designed using a specific set of Phase I samples obtained by practitioners. The $AARL_{\hat{c}|c_0}$ metric does not account for the variability in the in-control ARL values. The $AARL_{\hat{c}|c_0}$ value might be close to the targeted in-control ARL value, but the individual in-control $ARL_{\hat{c}|c_0}$ values can vary widely. The use of three-sigma limits can also mask the problem of practitioner-to-practitioner variability because charts with three-sigma limits effectively ignore the desired false alarm rate α.

Table IV. Percentage of *c*-charts with conditional ARL values below the targeted ARL values, that is, $ARL_{\hat{c}|c_0} < ARL_{c_0}$

	c_0			
m	3	10	20	50
20	42.74	38.19	32.92	64.30
50	36.08	24.53	14.71	49.33
100	29.99	13.61	5.52	34.03
1000	4.03	0.04	0.00	1.04
5000	0.00	0.00	0.00	0.00
ARL_{c_0}	262.95	160.68	102.86	114.58

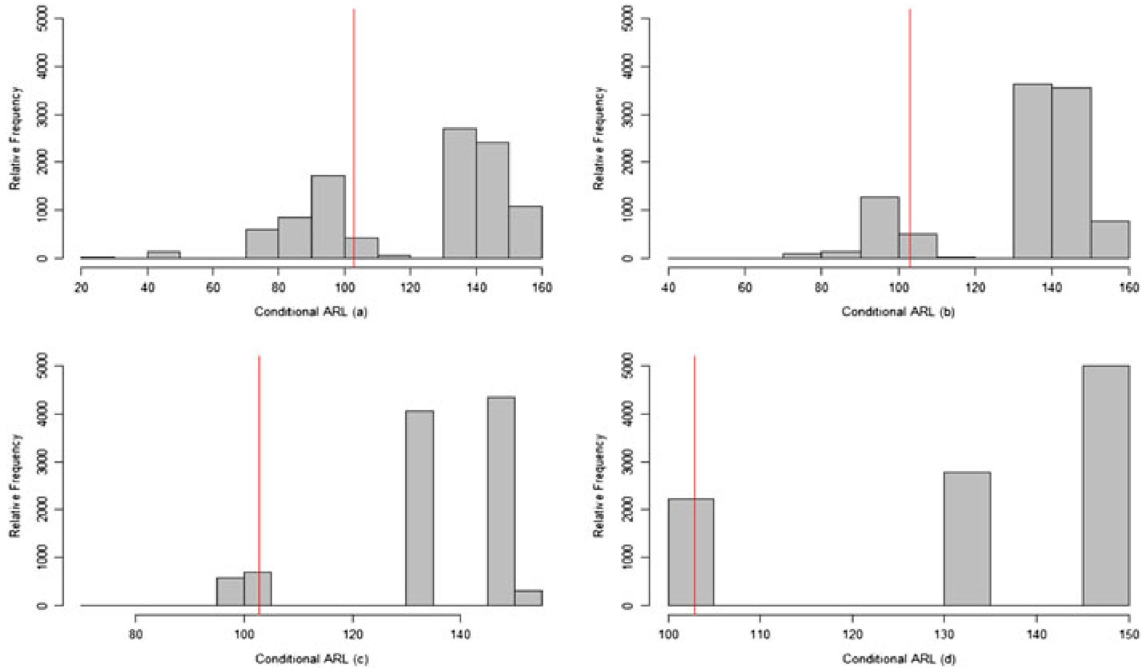

Figure 1. Histograms showing the distributions of the conditional in-control ARL values, $ARL_{\hat{c}|c_0}$, for $c_0 = 20$ for varying Phase I sample sizes. (a) $m = 20$; (b) $m = 50$; (c) $m = 100$; and (d) $m = 1000$. The red reference line in each plot refers to the targeted ARL, $ARL_{c_0} = 102.86$

Adjusted control limits using a bootstrap approach have been recommended to reduce the effect of sampling variability on the ARL performance of control charts with estimated parameters. The bootstrap approach taken by Jones and Steiner[16], Gandy and Kvaløy[17], and Saleh *et al.*[18] was used to adjust the control limits using bootstrapped samples from the Phase I data. From the bootstrapped samples, an approximate distribution of the standardized distances from the bootstrapped sample mean to the Phase I control limits is obtained. These distributions are used to adjust the control limits to improve the conditional in-control average run length performance of the control chart by controlling the proportion of charts with in-control ARL values below a given value. For the *c*-chart, we have adapted this bootstrapping approach. Instead of using the bootstrapped sample means to obtain a distribution of standardized distances, we use the bootstrap method to obtain a sampling distribution of the estimated mean $\hat{c}$ and then use percentiles from this distribution to adjust the control limits. This approach is considered because we are using probability limits, and these limits are solely based on the desired false alarm rate and the estimated or known mean.

In the computation of the bootstrapped control limits, first, the unadjusted control limits from the Phase I sample are calculated. Then a large number of bootstrapped samples each of size m are generated based on the estimated mean, $\hat{c}$. From the generated datasets, the distribution of the estimated means of the bootstrapped samples is saved onto a storage vector. Adjusted control limits are determined based on percentiles of this distribution. The adjustments widen the control limits to counteract the effect of estimation errors resulting from the unknown parameters. The bootstrap algorithm is outlined below.

1. Calculate c from your Phase I dataset where $\hat{c} = \frac{1}{m}\sum_{i=1}^{m} X_i$ for $i = 1, 2, \ldots, m$.
2. Calculate the unadjusted upper and lower control limits using Equation (6) or (8) if there is no lower control limit.
3. Generate $j = 1, \ldots, B$ bootstrap samples $X^*_{1,j}, X^*_{2,j}, \ldots, X^*_{m,j}$ from a $Poisson(m, \hat{c})$.
4. Calculate $\overline{b}_j = \frac{1}{m}\sum_{i=1}^{m} X^*_{i,j}$, the mean for each bootstrap sample.
5. Store the calculated $\overline{b}_j$ value for each bootstrap sample in a storage vector.
6. Find the pth and $(1-p)$th percentiles of the distribution of $\overline{b}$, denote as $\overline{b}_L$ and $\overline{b}_U$, respectively. For cases when there is no lower control limit, store only the $(1-p)$th percentile.
7. Calculate the adjusted control limits using Equation (10) when an upper and lower control limited are needed and Equation (11) when only the upper control limit is necessary.

$$P\left(X \geq UCL_{\overline{b}} | \overline{b}_U\right) \leq \frac{\alpha}{2} \text{and } P\left(X \leq LCL_{\overline{b}} | \overline{b}_L\right) \leq \frac{\alpha}{2}. \tag{10}$$

$$P\left(X \geq UCL_{\overline{b}} | \overline{b}_U\right) \leq \alpha. \tag{11}$$

The limits found in step 7 of the algorithm, $LCL_{\overline{b}}$ and $UCL_{\overline{b}}$, are the limits that the practitioner will use for his or her *c*-chart.

6. In-control ARL performance comparisons of the *c*-chart with and without bootstrap adjusted control limits

A simulation study was performed to evaluate the *c*-chart with bootstrap adjusted control limits. In the simulation, the conditional in-control ARL performance for *c*-charts with $c_0 = 3$, 10, 20, and 50 with Phase I samples of size $m = 20$, 50, 100, and 1000 was studied using 3000 simulations. In order to determine the effectiveness of the bootstrap adjusted control limits, a metric is needed to evaluate the in-control ARL performance for the control charted with adjusted limits but evaluated when c_0 is known. We will refer to this in-control ARL values as $ARL_{\bar{b}|c_0}$. Therefore, we can compare the in-control conditional ARL performance for unadjusted limits, $ARL_{\hat{c}|c_0}$, to the in-control conditional performance for charts with bootstrap adjusted limits, $ARL_{\bar{b}|c_0}$. The adjusted control limits are computed using the bootstrap procedure covered in Section 5 where the 5th and 95th percentiles of the distribution of bootstrap estimates are chosen for the upper and lower control limit adjustment calculations, respectively. For the case when $c_0 = 3$, only the upper 95th percentile was chosen for the upper control limit adjustment calculation. The desired α level is set to 0.10.

In Table V, the ARL performance for *c*-charts with $c_0 = 20$ and $c_0 = 50$ is compared for various sample sizes. The percentage of *c*-charts with ARL values below the targeted ARL is presented for both adjusted and unadjusted control limits. The percentages for the unadjusted control charts are equivalent to the percentages given in Table IV and are displayed in Table V for comparison purposes. When $c_0 = 20$, the ARL_0 value is 102.86. Without adjusted control limits, 33.92 % of the *c*-charts will result in $ARL_{\hat{c}|c_0}$ values below 102.86. When the limits are adjusted, the percentage drops to 0.4%. As the Phase I sample size *m* increases, the percentage of charts with conditional ARL performance values below ARL_{c_0} decreases in both the adjusted and unadjusted cases, which is expected. The adjustment reduces the percentage of charts with conditional in-control ARL performance, $ARL_{\bar{b}|c_0}$, below the targeted ARL_{c_0} values for all sample sizes except when $c_0 = 20$ and $m = 1000$. For that specific case, the charts with unadjusted limits have no in-control ARL values below the ARL_{c_0} value, meaning there is no need for adjustment. For the cases where $c_0 = 50$, similar patterns are observed.

The comparison of in-control conditional ARL performance for *c*-charts with adjusted and unadjusted control limits is illustrated in Figure 2 for the case where $c_0 = 20$ and $m = 20$ and in Figure 3 for $c_0 = 50$ and $m = 20$. The reference line in the

Table V. Comparison of the percentage of *c*-charts with conditional in-control ARL values below the targeted ARL_0 value for charts with adjusted and unadjusted control limits when $c_0 = 20$ and 50 and $m = 20$, 50, 100, and 1000

	$c_0 = 20$ $ARL_{c_0} = 102.86$		$c_0 = 50$ $ARL_{c_0} = 114.58$	
m	Adjusted	Unadjusted	Adjusted	Unadjusted
20	0.40	32.92	2.00	64.30
50	0.00	14.71	0.50	49.33
100	0.00	5.52	0.00	34.03
1000	0.00	0.00	0.00	1.04

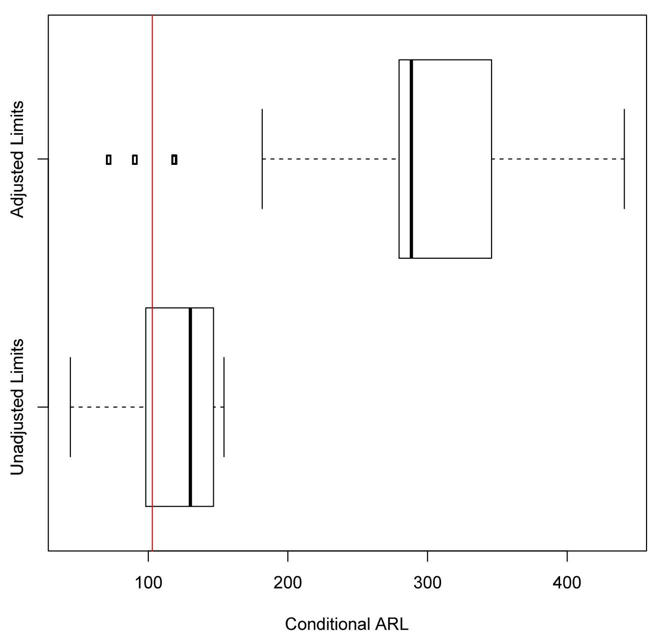

Figure 2. Boxplots showing in-control distribution for conditional ARL using adjusted and unadjusted control limits when $c_0 = 20$ and $m = 20$. The reference line is at ARL_{c_0} of 102.86

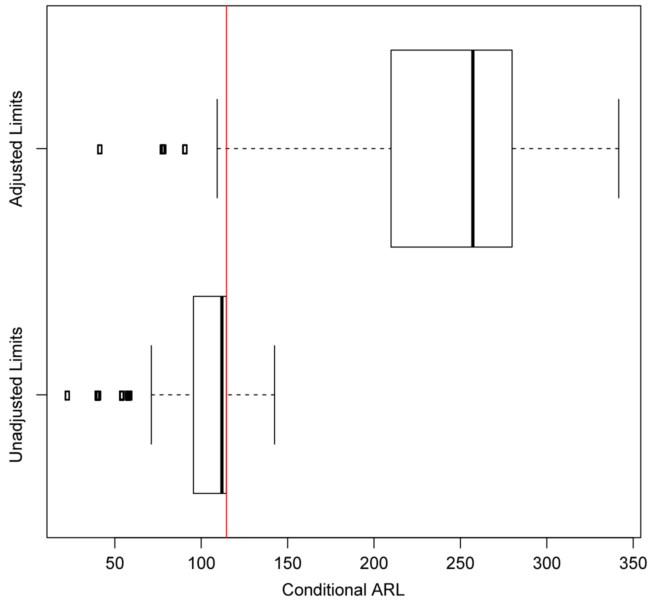

Figure 3. Boxplots showing in-control distribution for conditional ARL using adjusted and unadjusted control limits when $c_0 = 50$ and $m = 20$. The reference line is at ARL_{c_0} of 114.58

figures corresponds to the targeted ARL_{c_0} values when $c_0 = 20$ and $c_0 = 50$. These values are 102.86 for $c_0 = 20$ and 114.58 when $c_0 = 50$.

As the in-control number of non-conformities, c_0, decreases, there becomes less of a chance of having a lower control limit. In Table VI, the percentage of *c*-charts with in-control conditional ARL values below the targeted ARL_{c_0} value are compared for control charts with adjusted and unadjusted control limits when $c_0 = 3$ and $c_0 = 10$. In the case of $c_0 = 3$, there is no lower control limit, so only the upper control limit is adjusted using Equation (11). The bootstrap adjustment decreases the percentage of charts with in-control conditional ARL values below $ARL_{c_0} = 262.95$ for all Phase I sample sizes, *m*. The adjusted control limits also improve the conditional ARL performance in terms of the percentage of charts below $ARL_0 = 160.68$ for charts when $c_0 = 10$. Figures 4 and 5 illustrate the in-control conditional ARL performance for $c_0 = 3$ and $c_0 = 10$ when $m = 20$ for charts with both the adjusted and unadjusted control limits.

Overall, the parametric bootstrap approach has improved the integrity of the in-control ARL performance by decreasing the percentage of charts with ARL performance below the targeted ARL_{c_0} value. In Tables V and VI, it is interesting to note that with the parametric bootstrap adjustment, we are unable to guarantee a level of conditional ARL performance with a specified probability as is performed when the data is continuous. Again, this is a result of the discreteness of our data.

7. Out-of-control ARL performance comparisons of the *c*-chart with and without bootstrap adjusted control limits

When the control limits for the *c*-chart are adjusted to improve the in-control conditional ARL performance, this impacts the out-of-control ARL performance. A simulation study was performed to study this impact. The same bootstrapping procedure discussed in Section 5 was used to determine the control limits. This procedure was repeated 1000 times, and the out-of-control ARL performance was estimated for shifts in c_0 to a specified out-of-control mean c_1. The out-of-control performance was studied for the cases $c_0 = 3$ and $c_0 = 20$ when $m = 20$.

Table VI. Comparison of the percentage of *c*-charts with conditional in-control ARL values below the targeted ARL_0 value for charts with adjusted and unadjusted control limits when $\boldsymbol{c}_0 = 3$ and 10 and $m = 20$, 50, 100, and 1000

	$c_0 = 3$ $ARL_{c_0} = 262.95$		$c_0 = 10$ $ARL_{c_0} = 160.68$	
m	Adjusted	Unadjusted	Adjusted	Unadjusted
20	7.00	42.74	3.90	38.19
50	5.20	36.08	2.00	24.53
100	3.40	29.99	0.10	13.61
1000	0.10	4.03	0.00	0.04

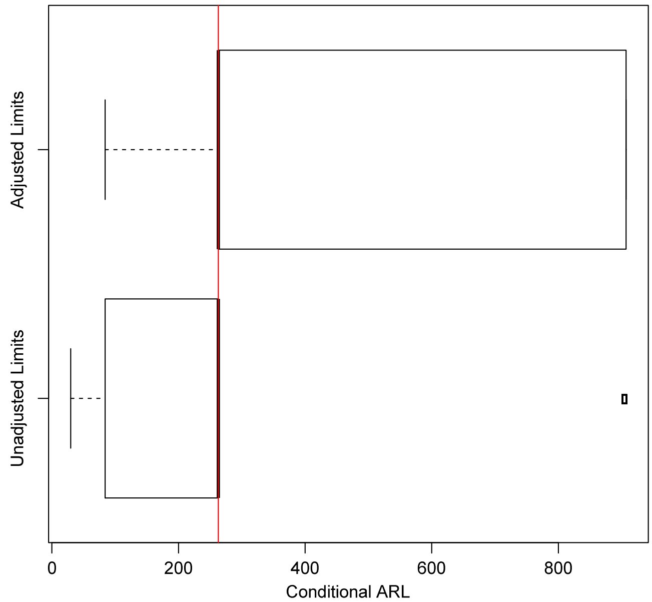

Figure 4. Boxplots showing In-control distribution for conditional ARL using adjusted and unadjusted control limits when $c_0 = 3$ and $m = 20$. The reference line is at ARL_{c_0} of 262.95

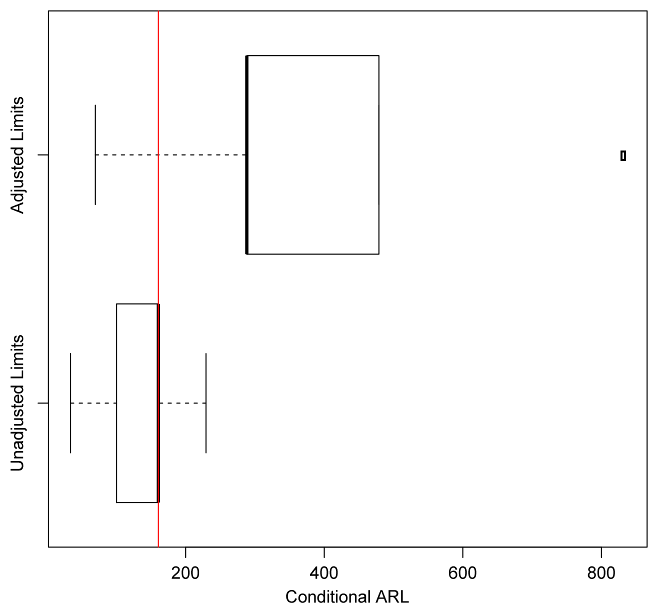

Figure 5. Boxplots showing in-control distribution for conditional *ARL* using adjusted and unadjusted control limits when c0 =10 and m = 20. The reference line is at ARL_{c_0} 160.68

The out-of-control conditional ARL values are compared for both *c*-charts with adjusted and unadjusted limits. Because the bootstrap method presented in this paper improves the run length of the *c*-chart when the process is in control by widening the control limits, one would expect some loss of sensitivity for detecting a process shift when using these new adjusted limits. The conditional out-of-control ARL performance for *c*-charts with both adjusted and unadjusted limits is illustrated in Figures 6 and 7 for $c_0 = 20$ and $c_0 = 3$, respectively. In both cases, the Phase I sample size $m = 20$. The shift is $c_1 = 22$ when $c_0 = 20$ and $c_1 = 4$ when $c_0 = 3$. There is more spread in the conditional out-of-control ARL performance for the control charts with adjusted limits in both scenarios. However, this reduction in sensitivity is necessary in order to ensure a certain level of in-control ARL performance.

The average out-of-control ARL performance for a range of shifts when $c_0 = 3$ and 20 for $m = 20$ is given in Table VII. For smaller shift sizes, adjusting the control limits lead to a more severe delay in the chart's detection of the out-of-control shift. However, as the shift becomes larger, the two charts have more similar detection performance.

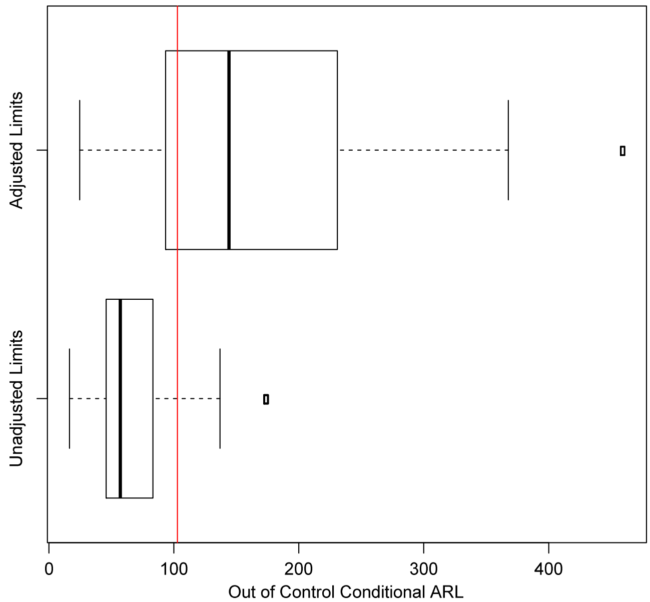

Figure 6. Out-of-control distribution for the conditional *ARL* using adjusted and unadjusted control limits when c_0 = 20, m = 20. The reference line is at ARL_{c_0} =102.86.

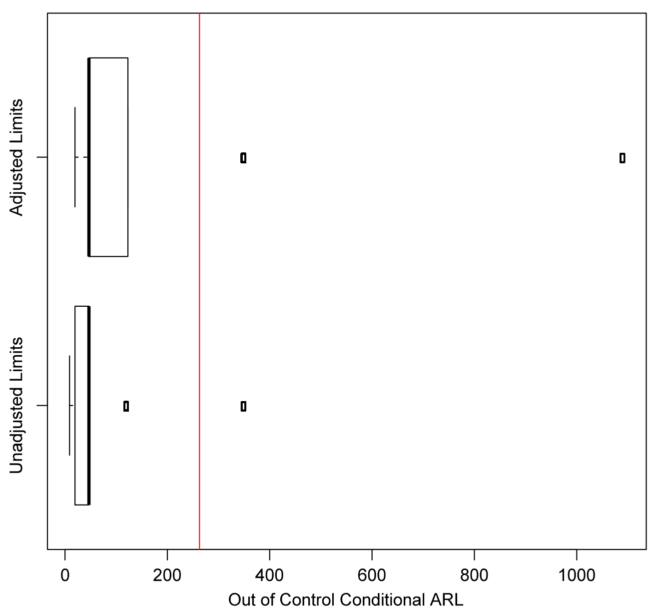

Figure 7. Out-of-control distribution for the conditional ARL using adjusted and unadjusted control limits when $c_0 = 3$, $m = 20$, and $c_1 = 4$. The reference line is at ARL_{c_0} = 262.95

Table VII. Out-of-control ARL performance when $c_0 = 3$ and $m = 20$

	$c_0 = 3$			$c_0 = 20$	
c_1	Adjusted	Unadjusted	c_1	Adjusted	Unadjusted
4	101.48	43.34	22	168.32	72.29
5	25.77	13.34	24	59.83	28.84
6	9.91	5.96	28	10.13	6.19
7	5.02	3.39	30	5.41	3.65

8. Conclusion

The conditional ARL performance of control charts must be taken into account when designing control charts that are based on estimated parameters from Phase I data. Previous research by Castagliola and Wu[9] and others has indicated that very large Phase I samples are necessary in order to overcome effects of estimation error. However, in their research, they used the conditional in-control ARL as the metric to determine the needed sample sizes. It is also important to consider the effect of estimation error on the conditional ARL performance. In this article, we propose a *c*-chart with adjusted probability limits to improve the conditional in-control ARL performance. We focused on probability limits for our *c*-chart. These limits are more appropriate than the common three-sigma limits especially when the distribution of the quality characteristic is not symmetric. The control limits are adjusted using a bootstrap approach as discussed in Jones and Steiner[16] and Gandy and Kvaløy[17]. The adjusted control charts result in improved in-control ARL performance. This was apparent by the reduction in the percentage of control charts with in-control conditional ARL values below the targeted ARL value. We recommend the use of the bootstrap adjusted control limits to improve in-control ARL performance. Because of the discreteness of the data, the bootstrap adjusted control limits do not always lead to conditional ARL performance with a specified probability, but for most cases, the conditional in-control ARL performance is drastically improved. Our research also suggests using probability limits as the upper and lower control limits for the *c*-chart. These probability limits lead to an inherent robustness in terms of in-control ARL performance because small deviations from the true mean c_0 that result in estimation will have no impact on the specified value of the control limits, leading to no change in in-control ARL performance.

In future research, it would be interesting to study ways of adapting the bootstrap approach for adjusting the control limits when the in-control number of non-conformities is small. It would also be of interest to explore alternative methods for determining the upper and lower control limits. Currently, we are determining the probability limits by splitting the desired false alarm rate, α, in half and ensuring that the chance of a false alarm on each side of the chart is no more than $\alpha/2$. There may be more advantageous ways for determining these cutoffs, such as a dynamic search algorithm. Another promising area of future research would be to study the conditional ARL performance for control charting where the assumed distribution follows difference discrete distributions.

References

1. Woodall WH. Control charts based on attribute data: bibliography and review. *Journal of Quality Technology* 1997; **29**(2):172–183.
2. Topalidou E, Psarakis S. Review of multinomial and multiattribute quality control charts. *Quality and Reliability Engineering International* 2009; **25**(7):773–804.
3. Chakraborti S, Human SW, Graham MA. Phase I statistical process control charts: an overview and some results. *Quality Engineering* 2009; **21**(1):52–62.
4. Jones-Farmer LA, Woodall WH, Steiner SH, Champ CW. An overview of phase I analysis for process improvement and monitoring. *Journal of Quality Technology* 2014; **46**(3):265–280.
5. Braun WJ. Run length distributions for estimated attributes charts. *Metrika* 1999; **50**(2):121–129.
6. Chakraborti S, Human SW. Parameter estimation and performance of the p-chart for attributes data. *IEEE Transactions on Reliability* 2006; **55**(3):559–566.
7. Chakraborti S, Human SW. Properties and performance of the c-chart for attributes data. *Journal of Applied Statistics* 2008; **35**(1):89–100.
8. Raubenheimer L, van der Merwe AJ. Bayesian control chart for nonconformities. *Quality and Reliability Engineering International* 2015; **31**(8):1359–1366.
9. Castagliola P, Wu S. Design of the C and np charts when the parameters are estimated. *International Journal of Reliability, Quality and Safety Engineering* 2012; **19**(02):1250010.
10. Castagliola P, Wu S, Khoo MBC, Chakraborti S. Synthetic phase II shewhart-type attributes control charts when process parameters are estimated. *Quality and Reliability Engineering International* 2014; **30**(3):315–335.
11. Testik MC. Conditional and marginal performance of the Poisson CUSUM control chart with parameter estimation. *International Journal of Production Research* 2007; **45**(23):5621–5638.
12. Jensen WA, Jones-Farmer LA, Champ CW, Woodall WH. Effects of parameter estimation on control chart properties: a literature review. *Journal of Quality Technology* 2006; **38**(4):349–364.
13. Psarakis S, Vyniou AK, Castagliola P. Some recent developments on the effects of parameter estimation on control charts. *Quality and Reliability Engineering International* 2014; **30**(8):1113–1129.
14. Saleh NA, Mahmoud MA, Keefe MJ, Woodall WH. The difficulty in designing Shewhart $\overline{X}$ and X control charts with estimated parameters. *Journal of Quality Technology* 2015; **47**(2):127–138.
15. Albers W, Kallenberg WCM. Are estimated control charts in control? *Statistics* 2004; **38**(1):67–79.
16. Jones MA, Steiner SH. Assessing the effect of estimation error on risk-adjusted CUSUM chart performance. *International Journal for Quality in Health Care* 2012; **24**(2):176–181.
17. Gandy A, Kvaloy JT. Guaranteed conditional performance of control charts via bootstrap methods. *Scandinavian Journal of Statistics* 2013; **40**(4):647–668.
18. Saleh NA, Mahmoud MA, Jones-Farmer LA, Zwetsloot I, Woodall WH. Another look at the EWMA control chart with estimated parameters. *Journal of Quality Technology* 2015; **47**(4):363–382.

Authors' biographies

Meng Zhao is currently a Ph.D. candidate in the Department of Statistics at Virginia Tech. He is expected to receive his doctorate in statistics in 2017. He received a M.S. in statistics in the Department of Statistics at Virginia Tech, and a M.S. in biotechnology at Pennsylvania State University. His Ph.D. research focuses in areas of statistical process monitoring. He is a student member of ASA.

Anne Driscoll is an Assistant Professor of Practice in the Department of Statistics at Virginia Tech. She received her PhD in Statistics from Virginia Tech. Her research interests include statistical process monitoring, design of experiments, and statistics education. She is a member of ASQ.