Research Article

On the Effect of Estimation Error for the Risk-Adjusted Charts

Sajid Ali ¹, **Naila Altaf**,¹ **Ismail Shah** ¹, **Lichen Wang** ²,
and **Syed Muhammad Muslim Raza** ³

¹Department of Statistics, Quaid-i-Azam University, Islamabad 45320, Pakistan

²College of Life Science, Linyi University, Linyi, Shandong 276000, China

³Department of Quantitative Methods, School of Business and Economics, University of Management and Technology, Lahore, Pakistan

Correspondence should be addressed to Lichen Wang; wanglichen@lyu.edu.cn

Received 12 July 2019; Accepted 24 December 2019; Published 17 February 2020

Academic Editor: Michele Scarpiniti

Copyright © 2020 Sajid Ali et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Control charts are a popular statistical process control (SPC) technique for monitoring to detect the unusual variations in different processes. Contrary to the classical charts, control charts have also been modified to include covariates using regression approaches. This study assesses the performance of risk-adjusted control charts under the complexity of estimation error by considering logistic and negative binomial regression models. To be more precise, risk-adjusted Cumulative Sum (CUSUM) and Exponentially Weighted Moving Average (EWMA) charts are used to evaluate the impact of the estimation error. To compute the average run length (ARL), Markov Chain Monte Carlo simulations are conducted. Furthermore, a bootstrap method is also used to compute the ARL assuming different Phase-I data sets to minimize the effect of estimation error on risk-adjusted control charts. The results for cardiac surgery and respiratory disease data sets show that the modified control charts improve the performance in detecting small shifts.

1. Introduction

The history of quality control is as old as the origin of industry. Statistical process control techniques can be applied to various kinds of information such as clinical outcomes, manufacturing processes, hazard management, and patient satisfaction; see, for example, Lowry and Montgomery [1], Woodall [2], Anderson and Thompson [3], Duncan [4], and Stoumbos et al. [5]. There is, however, a sharp distinction between industrial and healthcare applications. In industrial settings items are produced under very controlled conditions yielding mostly homogeneous products. Considering typical healthcare applications, on the other hand, patients have a wide variability in their personal profiles. This diversity, also known as risk factors in the literature, can have a substantial impact on the outcome and must not be ignored.

Control charts are used to detect any undesirable changes in the process. There are two main types of control charts: memoryless and memory-type control charts. The

memoryless charts are also known as the Shewhart charts, and their major drawback is the usage of current information while ignoring the history of the process. This disadvantage makes the Shewhart control chart quite insensitive to small shifts in the process. To overcome this problem, alternatively, the CUSUM [6] and EWMA [7] control charts are suggested to detect small shifts in the processes [8]. The CUSUM was developed on the basis of the total deviations of successive samples from the target value. Thus, each point plotted on the chart represents the sum of the deviations up to that particular point. The CUSUM has been proven to be efficient in detecting small shifts in the process.

Recently, accelerated interest has been shown in monitoring heterogeneous clinical processes. The goal is to monitor and control the outcome of a clinical process by taking into account the risk elements, i.e., patient covariates. The standard EWMA approach considers equal risk factors for all patients, which makes the method ineffective when monitoring healthcare outcomes. For

instance, an unexpected increase in the number of failures may be due to the treatment of numerous high risk patients and not because of alternate in the healthcare service. This would result in increased false alarms in the monitoring. Likewise, treating various low-risk cases yield a decreased wide variety of failures and may also result in an undetected deterioration of the service when using the traditional EWMA. Grigg and Spiegelhalter [9] have addressed these issues and developed various promising techniques to include risk factors in process monitoring and control of healthcare processes.

If the data are heterogeneous, the explainable variation present in the samples must be accounted, and in such cases, risk-adjusted control charts are suggested. A regression model is applied to the sampled data, which takes into account the explainable variation or covariates. Chen [10] introduced a set of methods for monitoring adverse outcome of the surveillance system. Later, Gallus et al. [11] proposed an optimal procedure for surveillance of congenital malformations and pointed out that the proposed procedure represents a clear improvement over the existing ones and is more efficient than the CUSUM scheme in certain conditions. Williams et al. [12] discussed an application of the CUSUM chart for monitoring surgical performance. Poloniecki et al. [13] proposed cumulative risk-adjusted mortality chart for detecting changes in the death rate. Steiner et al. [14] discussed the problem of monitoring outcome in pediatric cardiac surgery. Lovegrove et al. [15] described an alternative approach that accounts for an individual cardiac surgeon's case-mix by explicitly incorporating the inherent risk faced by patients due to a combination of factors relating to their age and the degree of disease they have. Steiner et al. [16] proposed a new CUSUM method for accounting preoperative risk or surgical failure through the use of the likelihood-based weighted scheme. Some more recent work on risk-adjusted charts can be seen in [17–27] and references cited therein.

Woodall [2] discussed industrial applications for practitioners which are similar to healthcare monitoring in public-health surveillance. Thor et al. [28] systematically reviewed the literature regarding how statistical process control has been applied to healthcare quality improvement and discussed the benefits, limitations, barriers, and facilitating factors related to such methods.

To account the estimation effect on the risk-adjusted charts, Jones and Steiner [29] studied the binary CUSUM using data on patients undergoing coronary artery bypass surgery and assessed mortality up to 30 days postsurgery. Later, Gandy and Kvaløy [30] suggested an adjustment to overcome the estimation error problems. Urbietta et al. [31] compared the CUSUM and EWMA control charts for the negative binomial distribution assuming the daily number of hospitalizations due to respiratory diseases for people over 65-years old in São Paulo city, Brazil. Höhle and Paul [32] discussed the control charts based on the Poisson and negative binomial distribution for monitoring time series of counts typically arising in the surveillance of infectious

diseases. More relevant discussion can be seen in [27, 33, 34].

The aim of this study is to discuss the problem of estimation error on the risk-adjusted logistic and negative binomial CUSUM charts. The performance of the charts is evaluated in terms of average run length (ARL). To account the estimation error, bootstrapping with different phases cardiac surgery data set is used.

The rest of the article is organized as follows. Section 2 defines the standard CUSUM chart and risk-adjusted CUSUM chart, whereas Section 3 presents the risk-adjusted EWMA chart. The negative binomial regression model is discussed in Section 4. The performance of the risk-adjusted CUSUM chart is presented in Section 5, whereas a bootstrap method to account estimation error is given in Section 6. The control limits assuming estimation error and different Phase-I data sets are studied in Section 7. The risk-adjusted EWMA performance is presented in Section 8. The analysis of respiratory disease patients and cardiac patients assuming negative binomial regression is discussed in Section 9. Section 10 presents some concluding remarks and proposed extensions.

2. Risk-Adjusted CUSUM Chart

The general test statistic of the tabular or standard CUSUM is described as

$$C_i = \max(0, C_{i-1} + W_i), \quad i = 1, 2, 3, \dots, \quad (1)$$

where $C_0 = 0$ and W_i is the sample weight assigned to the i th subgroup.

In surgical situations, pre-estimated mortality risk varies from patient to patient and the chart, given in equation (1), does not consider the preoperative risk factor. For this reason, an adjustment is required to ensure that sudden unusual changes are not due to surgeons. This can be done on the basis of prior risk by adopting a weighting scheme which allots higher values to higher risk patient and vice versa. The risk-adjusted (RA) CUSUM chart is used to detect a shift from the baseline rate to a specified alternative rate, where the shift is expressed in terms of odds ratio (OR). In general, OR for alternatives should be selected corresponding to the smallest change that is needed to be detected by the monitoring process. Let R_0 and R_1 represent the odds ratio under the null and the alternate hypotheses, respectively:

$$\begin{aligned} H_0: \text{odds ratio} &= R_0, \\ H_1: \text{odds ratio} &= R_1, \end{aligned} \quad (2)$$

where $R_0 = 1$ shows in-control state of the process and $R_1 > R_0$ refers to an increase in the odds ratio, while $R_0 > R_1$ indicates decrease. Thus, for patient i under H_0 , the odds of failure equals to $R_0 p_i / (1 - p_i)$ and its corresponding probability of failure equals to $R_0 p_i / (1 - p_i + R_0 p_i)$. The weighting score can be obtained by $p^{y_i} (1 - p)^{1 - y_i}$ for patient i , whereas the log-likelihood ratio score in the presence of the estimation error can be expressed as

$$W_i = \log \begin{cases} \left(\frac{1 + \Delta}{1 + (R1 - 1) * pi + \Delta} \right), & \text{if } y_i = 0, \\ \left(\frac{R1 + \Delta}{1 + (R1 - 1) * pi + \Delta} \right), & \text{if } y_i = 1, \end{cases} \quad (3)$$

where Δ represents the estimation error. The above weighted score is used in standard CUSUM equation (1), generally known as the risk-adjusted logistic CUSUM chart. If C_i is greater than control limit (h), the system shows a signal of out-of-control situation. Generally, the threshold h is chosen to maintain a specific false alarm rate.

The data set considered here is about the cardiac surgery collected by Cardiac Center of United Kingdom during 1992–1998 [16, 35]. There are 6,994 patients of cardiac surgery indicating 5,212 males and 1,782 females. The data set also contains information on age, gender, blood pressure, diabetic condition, and other related useful information indicating surgeon group, date, and risk factor which are stored as the Parsonnet score [35]. The average patient age is 62.50 with standard deviation 11. Further, after 30 days of surgery, 461 patients did not survive, i.e., the death rate is 6.6%.

2.1. Logistic Regression. For binary response variable Y_i , $i = 1, \dots, n$, the likelihood ratio is desired to construct a risk-adjusted chart. To find the likelihood ratio, the logistic regression model for the data must be obtained to estimate the logarithm of the odds. Mathematically, it can be expressed as

$$\text{logit}(P(Y_i = 1 | X_i)) = \ln\left(\frac{p_i}{1 - p_i}\right) = \beta_0 + X_i\beta_1, \quad (4)$$

where p_i is the likelihood that the i th sample is a success, such that $Y_i = 1$, X_i is the vector of input variables, and β is the coefficient for the input variables in the regression model [16]. Equation (4) can be rearranged to have the likelihood for $Y_i = 1$ as follows:

$$p_i = \left(\frac{\exp(\beta_0 + X_i\beta_1)}{1 + \exp(\beta_0 + X_i\beta_1)} \right). \quad (5)$$

Now, the variable Y_i follows the Bernoulli distribution with the probability mass function $P(y) = p^{y_i} (1 - p)^{1 - y_i}$, where $y_i \in \{0, 1\}$. The likelihood ratio (LR) then can be calculated by dividing the likelihood for the out-of-control situation, L_1 , and the likelihood for the in-control situation, L_0 . The L_0 can be calculated by

$$L_0 = p_0^{y_i} (1 - p_0)^{1 - y_i} = \frac{\exp(\beta_0 + X\beta_1)^{y_i}}{1 + \exp(\beta_0 + X\beta_1)}. \quad (6)$$

To observe the change in the logit model, the logit model for an out-of-control situation can be given as

$$\text{logit}(P(Y_i = 1 | X_i)) = \Delta + \beta_0 + X_i\beta_1, \quad (7)$$

where Δ denotes the out-of-control change in the logit function. Using equation (7), the possibility of an out-of-control situation can be calculated as

$$L_1 = p_1^{y_i} (1 - p_1)^{1 - y_i} = \frac{\exp(\Delta + \beta_0 + X\beta_1)^{y_i}}{1 + (\Delta + \exp(\beta_0 + X\beta_1))}. \quad (8)$$

Thus, the likelihood ratio can be expressed as

$$R = \log\left(\frac{p_1^{y_i} (1 - p_1)^{1 - y_i}}{p_0^{y_i} (1 - p_0)^{1 - y_i}}\right). \quad (9)$$

For patient i , the above equation can be written as

$$\exp(R_i) = \frac{\exp(\Delta)^{Y_i} \exp(1 + \beta_0 + X_i\beta_1)}{1 + (\Delta + \exp(\beta_0 + X_i\beta_1))}. \quad (10)$$

Now, for detecting change in the logit model, a RA-CUSUM chart based on the log-likelihood of the out-of-control verses in-control situation can be defined as

$$\begin{aligned} C_i &= \max(0, C_{i-1} + R_i), \\ C_i &= \max(0, C_{i-1} - R_i), \end{aligned} \quad (11)$$

where $C_0 = 0$. When the out-of-control situation is likely greater than the in-control situation, $R_i > 0$, it results in larger C_i values than the case of in-control situation. To monitor the increase in the odds for the out-of-control situation, Δ must be set greater than zero. Similarly, Δ should be less than zero to show a decrease in the odds.

3. RA-EWMA Charts

In SPC, the exponentially weighted moving average (EWMA) charts are very popular and widely used, like CUSUM charts, to detect small shifts. The monitoring statistic can be expressed as

$$C_i = \lambda R_i + (1 - \lambda)C_{i-1}, \quad (12)$$

where R_i is the EWMA score assigned to the i th observation and λ is a smoothing constant that lies between $0 < \lambda \leq 1$. Essentially, the statistic is a linear combination of all the previous observations with higher weights assigned to more previous observations than the current observations.

For detecting small changes, the RA-EWMA charts have a similar performance to CUSUM charts. However, the main advantage of RA-EWMA over CUSUM charts lies in its intuitive interpretation as the EWMA statistic can be seemed as an estimate of the current level of process. Moreover, by adjusting the weights rather than resetting the statistic as the CUSUM does, the influence of previous observations is removed gradually in the statistics which is a more natural way to conduct monitoring and is acceptable for healthcare practitioners [36, 37].

4. Negative Binomial Distribution Count Model

Most commonly used counts models are the Poisson and the negative binomial models. The probability density function of X_t following the Poisson distribution can be written as the member of the exponential family given below:

$$f(X_t | \mu_t) = \exp\{X_t \ln \mu_t - \mu_t - \ln(\Gamma(X_t + 1))\}, \quad (13)$$

where $E(X_t) = \text{Var}(X_t) = \mu_t$. By reparameterization of the above model for the negative binomial distribution, one can get the following model:

$$f(X_t | \mu_t, \phi) = \exp \left\{ X_t \ln \left(\frac{\mu_t}{\phi + \mu_t} \right) + \phi \ln \left(\frac{\phi}{\mu_t + \phi} \right) + \ln \left(\frac{\Gamma(\phi + X_t)}{\Gamma(X_t + 1)\Gamma(\phi)} \right) \right\}, \quad (14)$$

with $E(X_t) = \mu_t$, $\text{Var}(X_t) = \mu_t + (\mu_t^2/\phi)$ and Γ denotes the gamma function, i.e., $\Gamma(n) = (n-1)!$. Based on the generalized linear model (GLM) assuming the negative binomial distribution with log odds link function [38] and $\mu_{0,t}$ and $\mu_{1,t}$, as the in-control and out-of-control means, respectively, the CUSUM control chart can be written as

$$C_t = \max(0, C_{t-1} + S_t). \quad (15)$$

Thus, for the t th week under H_1 , the log odds link counts are $R_1\phi p_t/(1-p_t)$ and corresponding probability of counts equal to $R_1\phi p_t/(1-p_t + R_1\phi p_t)$. The weighting score is, thus, obtained by $\binom{y+\phi-1}{y} p^y (1-p)^\phi$. Finally, for t th week, the log-likelihood ratio score in the presence of estimation error can be expressed as

$$S_t = \log \begin{cases} \left[\frac{1 - (1 - R_0\phi)p_t}{1 - (1 - R_1\phi)p_t} \right]^\phi, & \text{if } y = 0, \\ \left[\frac{R_1(1 - p_t(1 - R_0\phi))}{R_0(1 - p_t(1 - R_1\phi))} \right]^{\phi+1}, & \text{if } y = 1. \end{cases} \quad (16)$$

The above weighted score is used to design risk-adjusted negative binomial CUSUM chart. If $C_t \geq h$ in equation (15), then chart shows a signal of out-of-control, which means that sufficient evidence has been collected that the underlying parameters are changed.

4.1. Negative Binomial Regression. A Poisson regression is used when there is no dispersion in the data, whereas the negative binomial regression is preferred for dispersed data. As discussed previously, for count response variables Y_t , $t = 1, \dots, n$, the likelihood ratio is used to obtain a risk-adjusted chart. To find the likelihood ratio for the negative binomial regression model, we proceed as follows:

$$P(Y = y) = \frac{\Gamma(\phi + y)}{\Gamma(y + 1)\Gamma(\phi)} \left(\frac{\phi}{\mu_t + \phi} \right)^\phi \left(\frac{\mu_t}{\phi + \mu_t} \right)^y, \quad (17)$$

$$\mu_t = \phi \left[(1 + \exp(x_t\beta))^{1/\phi} - 1 \right].$$

Now, the variable Y_t follows the negative binomial distribution with the probability mass function $P(y) = \binom{y+\phi-1}{y} p^y (1-p)^\phi$, where $y \in \{0, 1, 2, \dots\}$. The likelihood ratio (LR) is then calculated by dividing the

likelihood for the out-of-control situation, L_1 , and the likelihood for the in-control situation, L_0 , where L_0 can be calculated as

$$L_0 = \binom{y+\phi-1}{y} p_0^y (1-p_0)^\phi = \binom{y+\phi-1}{y} \left[(1 + \exp(\beta_0 + X\beta_1))^{1/\phi} - 1 \right]^y \left(\frac{1}{(1 + \exp(\beta_0 + X\beta_1))^{1/\phi}} \right)^{\phi+y}. \quad (18)$$

To observe the changes in the model, the model for an out-of-control situation is given by

$$\ln(\mu_t) = \Delta + \beta_0 + X_t\beta_1, \quad (19)$$

where Δ is the change in the log function. Using equation (19), the possibility of an out-of-control situation can be calculated as

$$L_1 = \binom{y+\phi-1}{y} p_1^y (1-p_1)^\phi = \binom{y+\phi-1}{y} \left[(1 + \exp(\Delta + \beta_0 + X\beta_1))^{1/\phi} - 1 \right]^y \cdot \left(\frac{1}{(1 + \exp(\Delta + \beta_0 + X\beta_1))^{1/\phi}} \right)^{\phi+y}. \quad (20)$$

The likelihood ratio is then written as

$$Q = \ln \left[\frac{\binom{y+\phi-1}{y} p_1^y (1-p_1)^\phi}{\binom{y+\phi-1}{y} p_0^y (1-p_0)^\phi} \right]. \quad (21)$$

For t th week, the above equation can be expressed as

$$\exp(Q) = \frac{[(1 + \exp(\Delta + x_t\beta))^{1/\phi} - 1][1/(1 + \exp(\Delta + x_t\beta))^{(\phi+y)/\phi}]}{[(1 + \exp(x_t\beta))^{1/\phi} - 1][1/(1 + \exp(x_t\beta))^{(\phi+y)/\phi}]}. \quad (22)$$

Finally, for detecting change in the logit model, a RA-CUSUM chart based on the log-likelihood of the out-of-control verses in-control situation can be defined as

$$C_i = \max(0, C_{i-1} + Q_i), \quad (23)$$

where $C_0 = 0$. A large value of C_t is observed in the out-of-control situation. To monitor the increase in the event for the out-of-control situation, Δ must be set greater than 0.

5. Performance of Risk-Adjusted CUSUM Charts

To evaluate the performance of control charts, Monte Carlo simulations and integral equation approaches are commonly used. In the class of Monte Carlo simulation, Markov Chain approach is a computationally efficient approach. To use the Markov Chain approach, the state space of the CUSUM, C_i ,

is partitioned into $m + 1$ states, i.e., CUSUM statistic lies between 0 and h , where h denotes the control limit by defining 0 state corresponding to the initial value and m state corresponding to absorbing state when $C_i > h$. The transition probability matrix for $m + 1$ possible states of CUSUM is given by

$$P = \begin{pmatrix} p_{00} & p_{01} & p_{02} & \cdots & p_{0m} \\ p_{10} & p_{11} & p_{12} & \cdots & p_{1m} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ p_{m0} & p_{m1} & p_{m2} & \cdots & p_{mm} \end{pmatrix}, \quad (24)$$

which can be rewritten as

$$P = \begin{pmatrix} \mathbf{R} & (\mathbf{I} - \mathbf{R}) \\ \mathbf{0} & \mathbf{1} \end{pmatrix}, \quad (25)$$

where

$$\mathbf{R} = \begin{pmatrix} p_{00} & p_{01} & p_{02} & \cdots & p_{0m-1} \\ p_{10} & p_{11} & p_{12} & \cdots & p_{1m-1} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ p_{m-10} & p_{m-11} & p_{m-12} & \cdots & p_{m-1m-1} \end{pmatrix}. \quad (26)$$

$\mathbf{I}$ is $m \times m$ identity matrix, $\mathbf{1}$ is $m \times 1$ column of ones, p_{ij} denotes the transition probability from state i to j , and $\mathbf{R}$ denotes the absorbing state without an out-of-control signal, i.e., just before the out-of-control situation. The last row and column of the matrix P shows an out-of-control signal corresponding to an absorbing state. If we remove this from the transition probability matrix, then the corresponding matrix is known as the $\mathbf{R}$ matrix. In general, it is suggested to set m approximately equal to 250 to have stable results. To elaborate it further, consider patients whose Parsonnet score is zero. Using equation (4), such low-risk patients have 2.5% estimated chance of death and 97.5% chance of survival, so from equation (3), assume $R_0 = 1$ and $R_1 = 2$, and the possible scores of such patients are 0.0669 and -0.024 .

Let ϕ represent the CUSUM run length, and then we can write

$$\begin{aligned} \Pr(\phi \leq i) &= (\mathbf{I} - \mathbf{R}^i)\mathbf{1}, \\ \Pr(\phi = i) &= (\mathbf{R}^{i-1} - \mathbf{R}^i)\mathbf{1}, \\ i &\geq 1. \end{aligned} \quad (27)$$

Thus, the expected value of run length is

$$E(\phi) = \sum_{i=1}^{\infty} i \Pr(\phi = i) = \sum_{i=1}^{\infty} (\mathbf{R}^{i-1} \mathbf{1}) = \sum_{i=1}^{\infty} (\mathbf{I} - \mathbf{R})^{-1}. \quad (28)$$

5.1. Estimation Error Impact. As mentioned previously, when the underlying distribution is unknown and chart parameters are to be estimated, the estimation error affects the efficiency of the control chart. To study the impact of estimated parameter(s) of the baseline distribution, the ARL is calculated through equation (3) using the Markov Chain

approach for different values of estimation error $\Delta \in 0(0.01)0.2$. The resulting study is listed in Tables 1–4.

The choice $R_1 = 2$ is used to show increasing the event rate, while $R_1 = 0.5$ for decreasing the event rate. From Tables 1 and 3, it is clear that as the estimation error increases, the corresponding ARL decreases. On the other hand, from Tables 2 and 4, the ARL increases when the estimation error increases using $R_1 = 0.5$. Hence, if R_1 is greater than one, the chart produces an out-of-control signal earlier than the case of R_1 less than one, suggesting that the process improvement cannot be detected in the presence of estimation. Hence, the estimation error has a significant effect on the chart performance.

6. Bootstrap Method

The best method for adjusting a monitoring scheme to account the estimation error is the use of the bootstrapping method proposed by Gandy and Kvaløy [30]. This section describes the bootstrapping procedure for obtaining the desired ARL on the basis of adjusted threshold.

Assuming the availability of stream of in-control independent observations, i.e., $X_1, X_2, \dots$ following a distribution P , a control chart is used to detect shifts when observations are no longer coming from the P , known as the out-of-control situation. To construct a chart, the parameters of P are estimated using the in-control data which are subject to the estimation error.

A common performance measure for control charts is the ARL which is defined as the hitting probability of the chart within a certain number of steps. To compute it, let τ denote the observation number at which the chart gives a signal of out-of-control, i.e., when the observation lies above the threshold h . The distribution of τ is also a function of estimated parameters. If τ is the unknown parameter, then we can express the ARL as $ARL(P; \xi) = E(\tau(\xi))$, where the expectation is with respect to P . The probability of signaling within m time steps can be expressed as $\text{hit}(P; \xi) = P(\tau(\xi) \leq m)$, for some finite $m > 0$. The threshold limit h is chosen to achieve a certain desired in-control performance of the chart. The chart signals when a threshold h is crossed, i.e., $h_{ARL}(P; \xi) = \inf\{h > 0: ARL(P; \xi) \geq \gamma\}$, where $\gamma > 0$. A threshold for the false alarm probability of β can be computed as $h_{hit}(P; \xi) = \inf\{h > 0: \text{hit}(P; \xi) \leq \beta\}$, i.e., $0 < \beta < 1$. Again, both, h_{ARL} and h_{hit} depend on P and ξ .

Let $\hat{P}$ and $\hat{\xi} = \xi(\hat{P})$ denote the estimated distribution and estimated parameters from the past in-control data $X_{-m}, \dots, X_{-1}$. In practice, $h_{ARL}(\hat{P}; \hat{\xi})$ or $h_{hit}(\hat{P}; \hat{\xi})$ is used, but it gives the performance far from the nominal level. To overcome this problem, Gandy and Kvaløy [30] suggested a bootstrap method to calculate the adjusted threshold, which with high probability guarantees the performance very close to the nominal value. To make a unified presentation of the bootstrap, let h be a common notation for quantities of interest like h_{ARL} and h_{hit} . Furthermore, let p_α be a constant such that $P[h(P; \hat{\xi}) - h(P; \tilde{\xi}) > p_\alpha] = 1 - \alpha$, which implies the following bound on the quantity of interest $P[h(P; \tilde{\xi}) < h(P; \hat{\xi}) - p_\alpha] = 1 - \alpha$.

TABLE 1: Impact of estimation error for $R_1 = 2$ and $h = 3$.

Δ	ARL
0	759.3629
0.01	467.9559
0.02	341.1733
0.03	274.6706
0.04	226.4678
0.05	182.2607
0.06	143.2288
0.07	121.6612
0.08	106.7489
0.09	94.3829
0.10	86.4473
0.11	80.2067
0.12	73.1504
0.13	63.6307
0.14	60.4502
0.15	56.2473
0.16	52.8653
0.17	48.4933
0.18	46.9376
0.19	44.7333
0.20	42.5858

TABLE 2: Impact of estimation error for $R_1 = 0.5$ and $h = 3$.

Δ	ARL
0	1177.207
0.01	1737.498
0.02	2743.781
0.03	4137.283
0.04	6219.137
0.05	9649.818
0.06	15114.42
0.07	22686.93
0.08	35074.18
0.09	54619.03
0.10	81145.06
0.11	370675.6
0.12	572213.2
0.13	900876.7
0.14	1425571
0.15	2221588
0.16	21756252
0.17	35310711
0.18	57533435
0.19	92415220
0.20	149063162

Since P is unknown, we cannot calculate p_α , but an approximation can be obtained by bootstrapping. Let $\hat{P}^*$ represent a parametric or nonparametric bootstrap to replicate the estimate of the in-control distribution $\hat{P}$ based on the sample size n and let $\hat{\xi}^* = \xi(\hat{P}^*)$. Then, we approximate p_α by p_α^* :

$$P\left[h(\hat{P}^*; \hat{\xi}^*) - h(\hat{P}; \hat{\xi}^*) > p_\alpha^* \mid \hat{P}\right] = 1 - \alpha. \quad (29)$$

Then, an approximate upper bound which guarantees a certain performance with an approximate probability of $(1 - \alpha)$ is $h(\hat{P}; \hat{\xi}) - p_\alpha^*$. Thus,

TABLE 3: Impact of estimation error for $R_1 = 2$ and $h = 5.03$.

Δ	ARL
0	10232.07
0.01	3026.098
0.02	1716.935
0.03	1209.824
0.04	544.5317
0.05	394.229
0.06	302.0776
0.07	228.2855
0.08	193.1768
0.09	176.3356
0.10	161.7523
0.11	142.553
0.12	127.5222
0.13	116.544
0.14	111.8552
0.15	102.8659
0.16	99.13175
0.17	92.28083
0.18	86.40121
0.19	81.93437
0.20	77.41752

TABLE 4: Impact of estimation error for $R_1 = 0.5$ and $h = 5.03$.

Δ	ARL
0	5381.892
0.01	10820.68
0.02	21931.29
0.03	71076.17
0.04	1415076
0.05	3385569
0.06	7909019
0.07	20677707
0.08	47530525
0.09	106786105
0.10	267728307
0.11	592242066
0.12	1282028980
0.13	2765284276
0.14	6607171082
0.15	13822972889
0.16	32267548395
0.17	66712000000
0.18	149423379512
0.19	306674508309
0.20	615605158640

$$(-\infty, h(\hat{P}; \hat{\xi}) - p_\alpha^*) \quad (30)$$

denotes the one-sided confidence interval for $(P; \hat{\xi})$. Algorithm 1 is a generic algorithm for bootstrapping.

Use $h = h_{\text{ARL}}$ or $h = h_{\text{hit}}$, and then run the chart with the adjusted control limit:

$$h(\hat{P}; \hat{\xi}) - p_\alpha^*. \quad (31)$$

6.1. Bootstrap in Risk-Adjusted Charts. Suppose we have pairs of independent observations $(Y_1, X_1), (Y_2, X_2), \dots$, where Y denotes response variable and X is the

From the past data $X_{-n}, \dots, X_{-1}$, estimate $\hat{P}$ and $\hat{\xi}$.

Generate bootstrap samples $X_{-n}^*, \dots, X_{-1}^*$ from $\hat{P}$. Compute the corresponding estimate $\hat{P}^*$ and $\hat{\xi}^*$. Repeat this step B times to get $\hat{P}_B^*, \dots, \hat{P}_1^*$ and $\hat{\xi}_B^*, \dots, \hat{\xi}_1^*$.

Let p_α^* be the $1-\alpha$ empirical quantile of $P(h(\hat{P}^*; \hat{\xi}^*) - h(\hat{P}; \hat{\xi}^*)), b = 1, \dots, B$.

ALGORITHM 1: Bootstrap.

corresponding covariate. Let P denote the joint distribution of (Y_i, X_i) and $\hat{P}$ be the empirical distribution which assigns weight $1/n$ to each past observation $(Y_{-n}, X_{-n}), \dots, (Y_{-1}, X_{-1})$. Then, by resampling from $\hat{P}$, the bootstrap can be applied to risk-adjusted charts. For example, consider the CUSUM chart with the in-control log odds ratio defined as $\text{logit}(P(Y_i = 1 | X_i)) = \beta_0 + X_i\beta_1$ and the out-of-control is $\text{logit}(P(Y_i = 1 | X_i)) = \Delta + \beta_0 + X_i\beta_1$. To observe changes in the odds ratio, the CUSUM is defined as

$$C_i = \max(0, C_{i-1} + R_i), \quad C_0 = 0, \quad (32)$$

where R_i is the log-likelihood ratio between the in-control and out-of-control model for observation i which gives a signal when $C_i > h$. In practice, β and the distribution of residuals P_ϵ are estimated, and then the CUSUM is run length with $h_{\text{ARL}}(\hat{P}_\epsilon; \hat{\beta})$.

The nonparametric bootstrap method can be used for accounting the estimation error by calculating the adjusted threshold, i.e., $\exp(\log h_{\text{ARL}}(\hat{P}_\epsilon; \hat{\beta}) - p_\alpha^*) = h_{\text{ARL}}(\hat{P}_\epsilon; \hat{\beta}) \exp(-p_\alpha^*)$. Using the adjusted threshold there is an approximate $(1 - \alpha)$ probability that the actual ARL of the chart is at least as large as the desired ARL.

7. Accounting Estimation Error in Cardiac Data

This section presents the results for the effect of Phase-I data on the RA-CUSUM.

Considering the first two-year data as the Phase-I data, we calculated the control limits using the procedure discussed in the previous section and the Phase-I data points are plotted against the constructed control limits using $R_1 = 2$. It is observed that out of 2,218 treated patients 143 died after 30 days of surgery. This gives probability of mortality $p = 0.06447$. Using $\text{logit}(p_i) = \beta_0 + X_i * \beta_1$ to detect $R_1 = 2$, the average in-control score is

$$\beta_0 + \bar{X}\beta_1 = \ln\left(\frac{p}{1-p}\right). \quad (33)$$

Now, substituting $p = 0.06447$, the average in-control value is

$$\beta_0 + \bar{X}\beta_1 = \ln\left(\frac{0.06447}{1-0.06447}\right) = -2.6749. \quad (34)$$

To observe the change, the out-of-control average risks score assuming the doubled mortality rate can be calculated as

$$\beta_0 + \bar{X}\beta_1 + \Delta = \ln\left(\frac{2 * p}{1 - 2 * p}\right) = \ln\left(\frac{2 * 0.06447}{1 - 2 * 0.06447}\right) = -1.9103. \quad (35)$$

Thus, substituting equation (35) in equation (34), we obtain

$$\Delta = -1.9103 + 2.6749 = 0.7646. \quad (36)$$

To obtain the adjusted control limits, the targeted in-control ARL is taken as 10,000, as the population is generally measured per ten thousand. The Phase-I and Phase-II risk-adjusted CUSUM charts are depicted in Figure 1 for the cardiac surgical data.

Figure 1(a) indicates the RA-CUSUM chart for the Phase-I observations and shows that for both, the unadjusted and the adjusted control limits, the Phase-I observations are in-control. Thus, we can use these limits for Phase-II data monitoring. Figure 1(b) indicates the RA-CUSUM chart for the Phase-II cardiac surgical operation data and show a steep increase in the mortalities before July 1995 and between July 1995 and January 1996; hence, the chart produces out-of-control signals, which indicates that patients are at risk and reason for this must be investigated. Using the unadjusted control limit chart, the first out-of-control observation is 3,729, whereas it is 3,830 for the adjusted limit chart. Thus, the false alarm rate decreases and efficiency of the chart increases by adjusting the control limit.

Next, considering the first three years as the Phase-I data, it is noticed that 3,187 patients are treated during this period, whereas 215 patients did not survive after 30 days of surgery; hence, the probability of mortality $p = 0.06746$. The average in-control score is

$$\beta_0 + \bar{X}\beta_1 = \ln\left(\frac{0.06746}{1-0.06746}\right) = -2.62635, \quad (37)$$

and the average out-of-control risks score can be calculated as

$$\beta_0 + \bar{X}\beta_1 + \Delta = \ln\left(\frac{2 * 0.06746}{1 - 2 * 0.06746}\right) = -1.85811. \quad (38)$$

Thus, the Δ can be calculated from equations (37) and (38) as follows:

$$\Delta = -1.85811 + 2.62635 = 0.768239. \quad (39)$$

The regression model becomes $\text{logit}(p_i) = -3.69043 + 0.07721 * X_i$. The Phase-I and Phase-II risk-adjusted CUSUM charts assuming first three years as the base period are depicted in Figure 2.

FIGURE 1: RA-CUSUM chart for Phase-I and Phase-II assuming first two years as the Phase-I for 30 days mortality of cardiac surgery patients. (a) Phase-I chart, unadjusted UCL=5.172 and adjusted UCL=6.225. (b) Phase-II chart, unadjusted UCL=5.172 and adjusted UCL=6.225.

FIGURE 2: RA-CUSUM chart for Phase-I and Phase-II assuming first three years as the Phase-I for 30 days mortality of cardiac surgery patients. (a) Phase-I chart, unadjusted UCL=5.184 and adjusted UCL=6.125. (b) Phase-II chart, unadjusted UCL=5.184 and adjusted UCL=6.125.

From Figure 2(b), it can be seen that the unadjusted control limit signals the first out-of-control signal at 3,729 position while the adjusted limit at 3,818. Hence, with the adjusted limit the false alarm rate decreases and charts

efficiency increases. Further, it is noticed that when the Phase-I data increases, i.e., from two years to three years, the adjusted control limit charts are more efficient to detect the change.

Next, we assess the performance of the RA-CUSUM to detect decrease in the odds occurrence for $R = 0.5$ assuming different Phase-I observations, as we did previously, to estimate the unknown parameters and construct control limits. The average out-of-control risks score in this case can be calculated as

$$\beta_0 + \bar{X}\beta_1 + \Delta = \ln\left(\frac{0.5 * 0.06447}{1 - 0.5 * 0.06447}\right) = -3.401896, \quad (40)$$

Thus, $\Delta = -3.401896 + 2.6749 = -0.727025$ and the regression model is $\text{logit}(p_i) = -3.6798 + 0.0768 * X_i$. The resulting study is depicted in Figure 3 for Phase-I and Phase-II risk-adjusted CUSUM charts.

Figure 3(b) shows the RA-CUSUM chart for the Phase-II cardiac surgical operation data. It can be seen that there is a steep increase before July 1997. By using the unadjusted limit, we received the first out-of-control signal at 5,233 position while at 5,355 for the adjusted limit chart. Thus, the adjusted chart produces less alarm than the unadjusted chart. Similarly, for first three-year data as the Phase-I, the average out-of-control risks score is

$$\beta_0 + \bar{X}\beta_1 + \Delta = \ln\left(\frac{0.5 * 0.06746}{1 - 0.5 * 0.06746}\right) = -3.355032, \quad (41)$$

and the value of Δ is

$$\Delta = -3.69043 + 2.62635 = -0.7286793. \quad (42)$$

This leads to the regression model $\text{logit}(p_i) = -3.69043 + 0.07721 * X_i$. The Phase-I and Phase-II risk-adjusted CUSUM charts assuming first three-year data as the Phase-I data are depicted in Figure 4.

Figure 4(b) shows that the first out-of-control signal appeared at 5,233 by the unadjusted chart while at 5,354 by the adjusted limit chart. Hence, the false alarm rate decreases by using the adjusted limit chart. Also, comparing charts constructed assuming two years Phase-I data to three years Phase-I data, that is, Figures 3 and 4, we get approximately very similar performance. In this case, the adjusted limit shows a signal of out-of-control at 5,355 observation by using two years data as the Phase-I while 5,354 by using three years data.

The Parsonnet scores are nonlinear on the logit scale, and to confirm this, we plot the histogram and normal probability plots in Figure 5.

Clearly, from the first row of the figure, the data do not follow the normal distribution. To transform the data to be approximately normal, the squared root transformation $\sqrt{X}$ is used and resulting study is plotted in the second row of Figure 5. Furthermore, the skewness of the original Parsonnet score is 1.9757 while kurtosis 8.4532. Similarly, the skewness and the kurtosis after the squared root transformation are 0.2475 and 2.9376, respectively. Thus, squared root transformation minimized the magnitude of skewness. Using the transformed data, the RA-CUSUM parameters are reestimated by using different Phase-I data sets to study the performance of chart in Phase-II. Considering the first two-year data as the Phase-I data, the Phase-I and Phase-II charts are depicted in Figure 6.

Figure 6(b) shows the risk-adjusted CUSUM chart for the Phase-II cardiac surgical operation data. From the figure it can be seen that, as discussed earlier, before July 1995 there is a steep increase in the mortalities. By using the unadjusted limit, the first out-of-control observation is 3,729, while with the adjusted limit, first out-of-control signal is observed at 3,901. Thus, by adjusting the limit, the false alarm rate decreases.

Next, considering the first three-year data as the Phase-I period, the graphical presentation of Phase-I and Phase-II data is given in Figure 7.

From Figure 7, it is observed that the first out-of-control signal is at 3,729 point by using unadjusted limit, while at 3,839 by using the adjusted limit chart. It is also noticed that by using a large Phase-I data, the adjusted limit chart is able to detect the change more quickly.

Furthermore, for $R_1 = 0.5$, assuming the first two-year data as the Phase-I data, the resulting charts are given in Figure 8. From Figure 8, it is noticed that by using the unadjusted chart the first out-of-control signal appeared at 5,142, whereas it is 5,227 using the adjusted chart. Similarly, using the first three-year data as the Phase-I data the graphical presentation of the data is depicted in Figure 9.

By the unadjusted chart, it is observed that the first out-of-control signal appears at 5,183, while it appears at 5,227 by the adjusted chart. Note that when we compare two years as the Phase-I data with three-year Phase-I, it is observed that the results of both charts are very similar, as both charts show out-of-control signal at 5,227. On comparing Parsonnet score results assuming transformed data by the squared root to nontransformed data, it is observed that the results of both regressions differ, and thus conclude that the identification of an appropriate distribution for data is the most important step in designing control charts. Furthermore, to supplement our conclusion, Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) of the models with original and transformed observations are calculated. The calculated values are 915 and 926.98 for untransformed data while 897 and 908.44 for the transformed data. Hence, the transformed data is more appropriate.

8. RA-EWMA Charts

As in the case of RA-CUSUM charts, here we adjust EWMA control limits to minimize the estimation error and to check the effect of Phase-I data.

Beginning with the first two-year data as the Phase-I data, we calculated the adjusted control limits by using smoothing constant $\lambda = 0.1$. Next, we used $R_1 = 2$ and plot the Phase-I and Phase-II data against the constructed control limits.

To obtain the adjusted control limit, 10,000 is assumed as the targeted in-control ARL. The regression model is then, $\text{logit}(p_i) = -3.68 + 0.077 * X_i$, where X_i is the Parsonnet score for the i th affected person. The Phase-I and Phase-II risk-adjusted EWMA charts are depicted in Figure 10 for the cardiac surgical data.

FIGURE 3: RA-CUSUM chart for Phase-I and Phase-II assuming first two years as the Phase-I for 30 days mortality of cardiac surgery patients. (a) Phase-I chart, unadjusted UCL = 5.093 and adjusted UCL = 5.97. (b) Phase-II chart, unadjusted UCL = 5.093 and adjusted UCL = 5.97.

FIGURE 4: RA-CUSUM chart for Phase-I and Phase-II assuming first three years as the Phase-I for 30 days mortality of cardiac surgery patients. (a) Phase-I chart, unadjusted UCL = 5.047 and adjusted UCL = 5.899. (b) Phase-II chart, unadjusted UCL = 5.047 and adjusted UCL = 5.899.

Figure 10(b) indicates the risk-adjusted EWMA chart for the Phase-II data and shows that since January 1996 there is a steep increase in the mortalities. By using the unadjusted and adjusted limit, the first out-of-control observation is

4,325, which indicates that adjusted limit has no significant effect on the EWMA chart.

Considering the first three-year data as the Phase-I period, the graphical presentation of Phase-I and Phase-II

FIGURE 5: Histogram and Normal Probability Plot of Parsonnet Score: first row-untransformed data and second row-transformed data. (a) Histogram. (b) Normal probability plot. (c) Histogram. (d) Normal probability plot.

data is given in Figure 11. From Figure 11(b), it is observed that the first out-of-control signal is at 4,325 point by using the unadjusted and adjusted limits. It is also noticed that despite using a large Phase-I data, the adjusted limit chart is unable to detect the change relatively quickly.

Next, we take first two-year data as the Phase-I data and calculate the adjusted control limits using smoothing constant $\lambda = 0.3$. Then, we use $R_1 = 2$ and plot the Phase-I and Phase-II data against the constructed control limits. The resultant Phase-I and Phase-II risk-adjusted EWMA charts are depicted in Figure 12. Similarly, letting the first three-year data as the Phase-I data, the graphical presentation of

Phase-I and Phase-II data is given in Figure 13. From Figures 12(b) and 13(b), note that all the point of Phase-II are in-control. Furthermore, by increasing λ value, the adjusted limit chart is unable to detect the small changes in the data.

Using the squared root transformation, the results assuming the first two-year data as the Phase-I data for different values of λ is presented in Figure 14. From the figure it is observed that by using two years data as the Phase-I data for different values of λ , i.e., (0.1 and 0.3), we get the same results as noticed previously in RA-EWMA charts. Similarly, from Figure 15 which corresponds to three-year Phase-I data for different values of λ , i.e., (0.1 and 0.3), we also get the

FIGURE 6: RA-CUSUM chart for Phase-I and Phase-II assuming first two years as the Phase-I for 30 days mortality of cardiac surgery patients. (a) Phase-I chart, unadjusted UCL = 5.238 and adjusted UCL = 6.522. (b) Phase-II chart, unadjusted UCL = 5.238 and adjusted UCL = 6.522.

FIGURE 7: RA-CUSUM chart for Phase-I and Phase-II assuming first three years as the Phase-I for 30 days mortality of cardiac surgery patients. (a) Phase-I chart, unadjusted UCL = 5.208 and adjusted UCL = 6.259. (b) Phase-II chart, unadjusted UCL = 5.208 and adjusted UCL = 6.259.

FIGURE 8: RA-CUSUM chart for Phase-I and Phase-II assuming first two years as the Phase-I for 30 days mortality of cardiac surgery patients. (a) Phase-I chart, unadjusted UCL = 4.57 and adjusted UCL = 5.575. (b) Phase-II chart, unadjusted UCL = 4.57 and adjusted UCL = 5.575.

FIGURE 9: RA-CUSUM chart for Phase-I and Phase-II assuming first three years as the Phase-I for 30 days mortality of cardiac surgery patients. (a) Phase-I chart, unadjusted UCL = 4.563 and adjusted UCL = 5.49. (b) Phase-II chart, unadjusted UCL = 4.563 and adjusted UCL = 5.49.

FIGURE 10: RA-EWMA chart for Phase-I and Phase-II assuming first two years as the Phase-I for 30 days mortality of cardiac surgery patients. (a) Phase-I chart, unadjusted UCL=0.19 and adjusted UCL=0.2006. (b) Phase-II chart, unadjusted UCL=0.19 and adjusted UCL=0.2006.

FIGURE 11: RA-EWMA chart for Phase-I and Phase-II assuming first three years as the Phase-I for 30 days mortality of cardiac surgery patients. (a) Phase-I chart, unadjusted UCL = 0.1918 and adjusted UCL = 0.2015. (b) Phase-II chart, unadjusted UCL = 0.1918 and adjusted UCL = 0.2015.

FIGURE 12: RA-EWMA chart for Phase-I and Phase-II assuming first two years as the Phase-I for 30 days mortality of cardiac surgery patients. (a) Phase-I chart, unadjusted UCL = 0.4281 and adjusted UCL = 0.4417. (b) Phase-II chart, unadjusted UCL = 0.4281 and adjusted UCL = 0.4417.

FIGURE 13: RA-EWMA chart for Phase-I and Phase-II assuming first three years as the Phase-I for 30 days mortality of cardiac surgery patients. (a) Phase-I chart, unadjusted UCL = 0.4314 and adjusted UCL = 0.4421. (b) Phase-II chart, unadjusted UCL = 0.4314 and adjusted UCL = 0.4421.

FIGURE 14: RA-EWMA chart for Phase-I and Phase-II assuming first two years as the Phase-I for 30 days mortality of cardiac surgery patients. (a) Phase-I chart, unadjusted UCL = 0.1871 and adjusted UCL = 0.198 and $\lambda = 0.1$. (b) Phase-II chart, unadjusted UCL = 0.1871 and adjusted UCL = 0.198 and $\lambda = 0.1$. (c) Phase-I chart, unadjusted UCL = 0.4213 and adjusted UCL = 0.4361 and $\lambda = 0.3$. (d) Phase-II chart, unadjusted UCL = 0.4213 and adjusted UCL = 0.4361 and $\lambda = 0.3$.

same results which we obtained previously in RA-EWMA charts. Thus, we conclude that both the CUSUM and the EWMA give the same results.

9. Negative Binomial Risk-Adjusted Chart for Cardiac Surgery Patient

This section fits GLM to analyze the count data of the weekly

hospital admission due to cardiac surgery in the UK. To this end, the logarithmic link function is used and population size is included as an offset to model the hospitalization rate per 100,000 inhabitants over time (Figure 16). The observed weekly admission of cardiac surgery patients along with the fitted model is depicted in Figure 17. In particular, the model includes sine and cosine functions to account for seasonal behavior while the week number to explain the variability of

FIGURE 15: RA-EWMA chart for Phase-I and Phase-II assuming first three years as the Phase-I for 30 days mortality of cardiac surgery patients. (a) Phase-I chart, unadjusted UCL = 0.189 and adjusted UCL = 0.1983 and $\lambda = 0.1$. (b) Phase-II chart, unadjusted UCL = 0.189 and adjusted UCL = 0.1983 and $\lambda = 0.1$. (c) Phase-I chart, unadjusted UCL = 0.424 and adjusted UCL = 0.4354 and $\lambda = 0.3$. (d) Phase-II chart, unadjusted UCL = 0.424 and adjusted UCL = 0.4354 and $\lambda = 0.3$.

FIGURE 16: Observed hospitalization rate and fitted model.

FIGURE 17: Weekly number of hospitalizations.

weekly hospitalizations. Let X_t denote the number of hospitalizations at week t with $X_t \sim \text{NegBin}(\mu_{0,t}, \phi)$, and then the model can be written as

$$\ln\left(\frac{\text{pop}_t}{\text{week}_t} 100,000\right) = \beta_0 + \beta_1 \text{week} + \sin\left(\frac{2\pi \text{week}}{52}\right) + \cos\left(\frac{2\pi \text{week}}{52}\right). \quad (43)$$

The inclusion of the offset term $g_t = (\text{week}_t/100000)$ allows to model the average weekly rate of admission in the hospital. The estimated coefficients of the model are presented in Table 5. The estimated dispersion parameter of the model is greater than one, indicating overdispersion.

TABLE 5: Model summary.

	Estimate	Std. error	z value	Pr(> z)
Intercept	-6.7856	0.0939	-72.27	0.0000
Week	0.0503	0.0033	15.04	0.0000
$\sin(2 * \text{pai} * \text{week}/52)$	0.1660	0.0732	2.27	0.0233
$\cos(2 * \text{pai} * \text{week}/52)$	-0.4461	0.0463	-9.63	0.0001
Dispersion	14.3795	3.19		

9.1. Effect of Estimation Error. To study the impact of estimated parameter of the baseline distribution, the ARL is calculated for different values of estimation error $\Delta \in 0(0.01)0.2$ and the resulting study is listed in Tables 6–9.

It is observed from the tables that the ARL decreases as the estimation error increases. Thus, the chart produces more signal as compared to the case when $\Delta = 0$, which shows the efficiency of the chart.

Next, to account for the estimation error, the bootstrap technique with different Phase-I data is used. It is observed that the first out-of-control point appeared at 246 week by the unadjusted chart ($\text{UCL} = 435.1$) assuming the first two years as the Phase-I data. However, using the adjusted limit ($\text{UCL} = 428.6$), the first out-of-control signal appears at 248 week. Similarly, for the first three-year weekly data as the Phase-I, the first out-of-control signal is detected at 298 week by the unadjusted limit ($\text{UCL} = 424.8$) while at 299 week for the adjusted limit ($\text{UCL} = 422.6$). Thus, using the adjusted and unadjusted control limits, no advantage is observed with increasing the Phase-I data.

9.2. Respiratory Disease Patient Analysis. This section presents the analysis of the weekly hospital admission due to respiratory diseases count data for people aged over 65 years in the city of São Paulo-Brazil [31]. Weekly data from January 2006 to December 2011 are used to fit the model, which consists of explanatory variable, such as week. Here, we consider negative binomial instead of Poisson because dispersion parameter ϕ is significantly larger than one. Additionally, the logarithmic odds link function is used and population size is included as an offset to model the hospitalization rate per 100,000 inhabitants over time. The weekly admission of the cardiac surgery patient is depicted in Figure 17. In the model, we included sine and cosine functions to account for seasonal behavior while week number to explain the variability of weekly hospitalizations. Let X_t denote the number of hospitalizations at week t with $X_t \sim \text{NegBin}(\mu_{0,t}, \phi)$, then the model can be written as

$$\ln\left(\frac{\mu_{0,t}}{\text{pop}_t} 100,000\right) = \beta_0 + \beta_1 \text{week} + \sin\left(\frac{2\pi \text{week}}{52}\right) + \cos\left(\frac{2\pi \text{week}}{52}\right). \quad (44)$$

The inclusion of the offset term $g_t = (\text{pop}_t/100000)$ allows to model the average weekly rate of admission in the hospital. The estimated coefficients of the model are presented in Table 10.

TABLE 6: Impact of estimation error for $R_1 = 2$, $\phi = 1$, and $h = 3$.

Δ	ARL
0	882.6903
0.01	436.3778
0.02	264.6784
0.03	183.7653
0.04	137.4939
0.05	109.3725
0.06	88.9749
0.07	76.6676
0.08	67.3711
0.09	60.2734
0.10	54.2372
0.11	49.6043
0.12	45.1277
0.13	42.0064
0.14	39.3413
0.15	37.0458
0.16	34.8897
0.17	33.1124
0.18	31.3214
0.19	38.0067
0.20	28.7036

TABLE 7: Impact of estimation error for $R_1 = 2$, $\phi = 1$, and $h = 5.03$.

Δ	ARL
0	5477.165
0.01	1533.957
0.02	651.7671
0.03	387.7923
0.04	268.0826
0.05	193.2517
0.06	158.0061
0.07	134.1011
0.08	115.3415
0.09	102.404
0.10	91.31192
0.11	81.14096
0.12	74.84708
0.13	69.41572
0.14	64.95874
0.15	61.04556
0.16	57.5161
0.17	54.18813
0.18	51.63691
0.19	49.33849
0.20	47.15687

TABLE 8: Impact of estimation error for $R_1 = 2$, $\phi = 2$, and $h = 3$.

Δ	ARL
0	53.7376
0.01	50.1265
0.02	46.6354
0.03	43.5539
0.04	41.1191
0.05	38.8163
0.06	36.9271
0.07	35.1309
0.08	33.4988
0.09	32.0334
0.10	30.741
0.11	29.5503
0.12	28.5108
0.13	27.5034
0.14	26.5793
0.15	25.7226
0.16	24.9543
0.17	24.2231
0.18	23.6096
0.19	22.9885
0.20	22.4013

TABLE 9: Impact of estimation error for $R_1 = 2$, $\phi = 2$, and $h = 5.03$.

Δ	ARL
0	90.762
0.01	83.3486
0.02	77.2786
0.03	72.2559
0.04	67.8309
0.05	64.2946
0.06	60.9486
0.07	57.9683
0.08	55.3429
0.09	52.9487
0.10	50.8527
0.11	49.0226
0.12	47.2249
0.13	45.4612
0.14	43.817
0.15	42.4548
0.16	41.2281
0.17	40.1236
0.18	39.0251
0.19	38.0067
0.20	37.055

9.3. Effect of Estimation Error. To study the impact of the estimated parameter of the baseline distribution, we calculated the ARL for different values of estimation error $\Delta \in 0(0.01)0.2$. The resulting study is listed in Tables 11 and 12.

From the tables, it is observed that the ARL decreases as the estimation error increases. Thus, the chart produces more out-of-control signals as compared to the case when $\Delta = 0$.

To account for the estimation error and to assess the effect of different Phase-I data, we considered first two years and then three years as the Phase-I data and

calculated the control limits. Using two year weekly data as the Phase-I, the first out-of-control point is observed at 250 week by the unadjusted chart ($UCL = 759.4$). However, using the adjusted limit ($UCL = 764.1$), the first out-of-control signal appears at 251 week. Similarly, using first three years weekly data as the Phase-I, the first point goes out-of-control at 303 week by the unadjusted limit ($UCL = 756.3$) while at 304 week for the adjusted limit ($UCL = 760.2$). Thus, we conclude that some improvement is observed by the adjusting control chart over the unadjusted chart to detect out-of-control point for this data.

TABLE 10: Model summary.

	Estimate	Std. error	z value	Pr(> z)
Intercept	-1.0062	0.0613	-16.42	0.0000
Week	-0.0006	0.0022	-0.27	0.7883
$\sin(2 * \text{pai} * \text{week}/52)$	-0.0484	0.0458	-1.06	0.2905
$\cos(2 * \text{pai} * \text{week}/52)$	-0.3892	0.0288	-13.52	0.0000
Dispersion	26.75	4.19		

TABLE 11: Impact of estimation error for $R_1 = 2$, $r = 1$, and $h = 5.32$.

Δ	ARL
0	501.698
0.01	366.1161
0.02	299.6614
0.03	251.2851
0.04	209.9261
0.05	172.1565
0.06	144.0628
0.07	128.8339
0.08	116.9328
0.09	106.7147
0.10	93.93775
0.11	83.56706
0.12	76.24845
0.13	71.2077
0.14	67.27537
0.15	63.98711
0.16	59.70038
0.17	55.95168
0.18	52.98182
0.19	51.0446
0.20	49.27074

TABLE 12: Impact of estimation error for $r = 2$ and $h = 78.60$.

Δ	ARL
0	500.0398
0.01	497.6541
0.02	494.4152
0.03	490.8751
0.04	491.2513
0.05	495.1931
0.06	498.015
0.07	494.421
0.08	492.2429
0.09	488.7528
0.10	485.2901
0.11	484.3075
0.12	487.2432
0.13	490.3458
0.14	487.2736
0.15	484.4512
0.16	481.4045
0.17	478.1174
0.18	474.5932
0.19	472.2021
0.20	474.1996

10. Conclusion

The main aim of this study is to assess the impact of the estimation error on the risk-adjusted control charts. In practice, the estimation error occurs when the distribution of the underlying data is unknown and we estimate the parameters from the Phase-I data. To account for this problem, a bootstrap method with different Phase-I data is used to establish adjusted limits of control charts which reduced the false alarm rate. In addition, by increasing Phase-I data, it is able to detect changes quickly. In particular, we considered risk-adjusted CUSUM charts assuming Bernoulli and negative binomial models. Further, risk-adjusted EWMA is also discussed in the study. As literature suggested, the bootstrapping methods can be computationally costly; however, we found it the most efficient method for minimizing the impact of the estimation error with low computational cost due to modern computing power. In future, other transformations, such as logarithmic and Box-Cox, can be considered to compare the performance of risk-adjusted CUSUM and EWMA charts.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

- [1] C. A. Lowry and D. C. Montgomery, "A review of multivariate control charts," *IIE Transactions*, vol. 27, no. 6, pp. 800–810, 1995.
- [2] W. H. Woodall, "The use of control charts in health-care and public-health surveillance," *Journal of Quality Technology*, vol. 38, no. 2, pp. 89–104, 2006.
- [3] M. J. Anderson and A. A. Thompson, "Multivariate control charts for ecological and environmental monitoring," *Ecological Applications*, vol. 14, no. 6, pp. 1921–1935, 2004.
- [4] A. J. Duncan, *Quality Control and Industrial Statistics*, Richard D. Irwin Inc., Homewood, IL, USA, 1974.
- [5] Z. G. Stoumbos, M. R. Reynolds Jr, T. P. Ryan, and W. H. Woodall, "The state of statistical process control as we proceed into the 21st century," *Journal of the American Statistical Association*, vol. 95, no. 451, pp. 992–998, 2000.
- [6] E. S. Page, "Continuous inspection schemes," *Biometrika*, vol. 41, no. 1/2, pp. 100–115, 1954.
- [7] S. W. Roberts, "Control chart tests based on geometric moving averages," *Technometrics*, vol. 1, no. 3, pp. 239–250, 1959.
- [8] D. C. Montgomery, *Introduction to Statistical Quality Control*, John Wiley & Sons, New York, NY, USA, 2009.
- [9] O. Grigg and D. Spiegelhalter, "A simple risk-adjusted exponentially weighted moving average," *Journal of the American Statistical Association*, vol. 102, no. 477, pp. 140–152, 2007.

- [10] R. Chen, "A surveillance system for congenital malformations," *Journal of the American Statistical Association*, vol. 73, no. 362, pp. 323–327, 1978.
- [11] G. Gallus, C. Mandelli, M. Marchi, and G. Radaelli, "On surveillance methods for congenital malformations," *Statistics in Medicine*, vol. 5, no. 6, pp. 565–571, 1986.
- [12] S. M. Williams, B. R. Parry, and M. M. Schlup, "Quality control: an application of the cusum," *BMJ*, vol. 304, no. 6838, pp. 1359–1361, 1992.
- [13] J. Poloniecki, O. Valencia, and P. Littlejohns, "Cumulative risk adjusted mortality chart for detecting changes in death rate: observational study of heart surgery," *BMJ*, vol. 316, no. 7146, pp. 1697–1700, 1998.
- [14] S. H. Steiner, R. J. Cook, and V. T. Farewell, "Monitoring paired binary surgical outcomes using cumulative sum charts," *Statistics in Medicine*, vol. 18, no. 1, pp. 69–86, 1999.
- [15] J. Lovegrove, C. Sherlaw-Johnson, O. Valencia, T. Treasure, and S. Gallivan, "Monitoring the performance of cardiac surgeons," *The Journal of the Operational Research Society*, vol. 50, no. 7, pp. 684–689, 1999.
- [16] S. H. Steiner, R. J. Cook, V. T. Farewell, and T. Treasure, "Monitoring surgical performance using risk-adjusted cumulative sum charts," *Biostatistics*, vol. 1, no. 4, pp. 441–452, 2000.
- [17] O. A. Grigg, V. T. Farewell, and D. J. Spiegelhalter, "Use of risk-adjusted CUSUM and RSPRT charts for monitoring in medical contexts," *Statistical Methods in Medical Research*, vol. 12, no. 2, pp. 147–170, 2003.
- [18] O. A. Grigg and V. T. Farewell, "A risk-adjusted sets method for monitoring adverse medical outcomes," *Statistics in Medicine*, vol. 23, no. 10, pp. 1593–1602, 2004.
- [19] O. Grigg and V. Farewell, "An overview of risk-adjusted charts," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, vol. 167, no. 3, pp. 523–539, 2004.
- [20] E. Gombay, A. A. Hussein, and S. H. Steiner, "Monitoring binary outcomes using risk-adjusted charts: a comparative study," *Statistics in Medicine*, vol. 30, no. 23, pp. 2815–2826, 2011.
- [21] J. Neuburger, K. Walker, C. Sherlaw-Johnson, J. van der Meulen, and D. A. Cromwell, "Comparison of control charts for monitoring clinical performance using binary data," *BMJ Quality & Safety*, vol. 26, no. 11, pp. 919–928, 2017.
- [22] C. D. Roberts, "Unbiased estimation of odds ratio by using negative binomial plans," *Biometrical Journal*, vol. 35, no. 5, pp. 581–587, 1993.
- [23] F. Mohammadian, S. T. A. Niaki, and A. Amiri, "Phase-I risk-adjusted geometric control charts to monitor health-care systems," *Quality and Reliability Engineering International*, vol. 32, no. 1, pp. 19–28, 2016.
- [24] A. P. Alencar, L. L. Ho, and O. Y. E. Albarracin, "CUSUM control charts to monitor series of negative binomial count data," *Statistical Methods in Medical Research*, vol. 26, no. 4, pp. 1925–1935, 2017.
- [25] G. Rossi, L. Lampugnani, and M. Marchi, "An approximate CUSUM procedure for surveillance of health events," *Statistics in Medicine*, vol. 18, no. 16, pp. 2111–2122, 1999.
- [26] S. Unkel, C. P. Farrington, P. H. Garthwaite, C. Robertson, and N. Andrews, "Statistical methods for the prospective detection of infectious disease outbreaks: a review," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, vol. 175, no. 1, pp. 49–82, 2012.
- [27] A. Sachlas, S. Bersimis, and S. Psarakis, "Risk-adjusted control charts: theory, methods, and applications in health," *Statistics in Biosciences*, vol. 11, no. 3, pp. 630–658, 2019.
- [28] J. Thor, J. Lundberg, J. Ask et al., "Application of statistical process control in healthcare improvement: systematic review," *Quality and Safety in Health Care*, vol. 16, no. 5, pp. 387–399, 2007.
- [29] M. A. Jones and S. H. Steiner, "Assessing the effect of estimation error on risk-adjusted CUSUM chart performance," *International Journal for Quality in Health Care*, vol. 24, no. 2, pp. 176–181, 2011.
- [30] A. Gandy and J. T. Kvaløy, "Guaranteed conditional performance of control charts via bootstrap methods," *Scandinavian Journal of Statistics*, vol. 40, no. 4, pp. 647–668, 2013.
- [31] P. Urbietia, L. L. Ho, and A. Alencar, "CUSUM and EWMA control charts for negative binomial distribution," *Quality and Reliability Engineering International*, vol. 33, no. 4, pp. 793–801, 2017.
- [32] M. Höhle and M. Paul, "Count data regression charts for the monitoring of surveillance time series," *Computational Statistics & Data Analysis*, vol. 52, no. 9, pp. 4357–4368, 2008.
- [33] L. A. Jones, C. W. Champ, and S. E. Rigdon, "The run length distribution of the CUSUM with estimated parameters," *Journal of Quality Technology*, vol. 36, no. 1, pp. 95–108, 2004.
- [34] M. Zhang, Y. Xu, Z. He, and X. Hou, "The effect of estimation error on risk-adjusted survival time CUSUM chart performance," *Quality and Reliability Engineering International*, vol. 32, no. 4, pp. 1445–1452, 2016.
- [35] A. Gandy and J. T. Kvaløy, "Spcadjust: functions for calibrating control charts," 2016, <https://cran.r-project.org/web/packages/spcadjust/index.html>.
- [36] D. A. Cook, M. Coory, and R. A. Webster, "Exponentially weighted moving average charts to compare observed and expected values for monitoring risk-adjusted hospital indicators," *BMJ Quality & Safety*, vol. 20, no. 6, pp. 469–474, 2011.
- [37] D. A. Cook, G. Duke, G. K. Hart, D. Pilcher, and D. Mullany, "Review of the application of risk-adjusted charts to analyse mortality outcomes in critical care," *Critical Care and Resuscitation*, vol. 10, no. 3, pp. 239–251, 2008.
- [38] C. J. Sroka and H. N. Nagaraja, "Odds ratios from logistic, geometric, Poisson, and negative binomial regression models," *BMC Medical Research Methodology*, vol. 18, no. 1, p. 112, 2018.