Dynamic Probability Control Limits for Risk-Adjusted Bernoulli Cumulative Sum Charts

Xiang Zhang

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
In
Statistics

William H. Woodall, Committee Chair
Anne G. Ryan Driscoll
Christopher T. Franck
Yili Hong

December 4th, 2015
Blacksburg, VA

Keywords: statistical process control, average run length (ARL), false alarm rate, run length distribution, surgical outcome quality.

Dynamic Probability Control Limits for Risk-Adjusted Bernoulli Cumulative Sum Charts

Xiang Zhang

ABSTRACT

The risk-adjusted Bernoulli cumulative sum (CUSUM) chart developed by Steiner *et al*. (2000) is an increasingly popular tool for monitoring clinical and surgical performance. In practice, however, use of a fixed control limit for the chart leads to quite variable in-control average run length (ARL) performance for patient populations with different risk score distributions. To overcome this problem, the simulation-based dynamic probability control limits (DPCLs) patient-by-patient for the risk-adjusted Bernoulli CUSUM charts is determined in this study. By maintaining the probability of a false alarm at a constant level conditional on no false alarm for previous observations, the risk-adjusted CUSUM charts with DPCLs have consistent in-control performance at the desired level with approximately geometrically distributed run lengths. Simulation results demonstrate that the proposed method does not rely on any information or assumptions about the patients' risk distributions. The use of DPCLs for risk-adjusted Bernoulli CUSUM charts allows each chart to be designed for the corresponding particular sequence of patients for a surgeon or hospital. The effect of estimation error on performance of risk-adjusted Bernoulli CUSUM chart with DPCLs is also examined. Our simulation results show that the in-control performance of risk-adjusted Bernoulli CUSUM chart with DPCLs is affected by the estimation error. The most influential factors are the specified desired in-control average run length, the Phase I sample size and the overall adverse event rate.

However, the effect of estimation error is uniformly smaller for the risk-adjusted Bernoulli CUSUM chart with DPCLs than for the corresponding chart with a constant control limit under various realistic scenarios. In addition, there is a substantial reduction in the standard deviation of the in-control run length when DPCLs are used. Therefore, use of DPCLs has yet another advantage when designing a risk-adjusted Bernoulli CUSUM chart. These researches are results of joint work with Dr. William H. Woodall (Department of Statistics, Virginia Tech). Moreover, DPCLs are adapted to design the risk-adjusted CUSUM charts for multiresponses developed by Tang *et al*. (2015). It is shown that the in-control performance of the charts with DPCLs can be controlled for different patient populations because these limits are determined for each specific sequence of patients. Thus, the risk-adjusted CUSUM chart for multiresponses with DPCLs is more practical and should be applied to effectively monitor surgical performance by hospitals and healthcare practitioners. This research is a result of joint work with Dr. William H. Woodall (Department of Statistics, Virginia Tech) and Mr. Justin Loda (Department of Statistics, Virginia Tech).

Dedication

To my parents and my advisor

Acknowledgements

First and foremost, I would like to sincerely and gratefully thank my advisor, Dr. William H. Woodall, for his guidance, support, patience, understanding and encouragement throughout the entire period of my graduate studies at Virginia Tech. He has helped me to become more independent, more optimistic, and most importantly, more confident about my academic career and has taught me to always have a "can-do" attitude towards various scenarios. For everything you have done for me, Dr. Woodall, I thank you.

I would like to thank the remaining members of my graduate committee, Dr. A Anne G. Ryan Driscoll, Dr. Christopher T. Franck and Dr. Yili Hong. Your consideration, suggestion and assistance are greatly appreciated. I would also like to express my gratitude to all the faculty, stuff and graduate students in Department of Statistics at Virginia Tech, who throughout my educational career have supported and encouraged me to believe in my abilities. They have guided me through many difficult situations, allowing me to reach this accomplishment. Special thanks to my co-authors, Wenmeng, Hongyue, Justin and my friends, Xiaoning, Peng, Matt *et al.* for their input and contributions to my papers and researches.

Finally, I would like to thank my family, especially my parents, for their firm belief in me. It would be impossible for me to achieve any goals without their unconditional love and support. Thank you so much, Mom and Dad. I am so grateful to be your son and will keep doing my best to make you proud.

Table of contents

List of figures

List of tables

Chapter 1

Introduction

Healthcare quality has been an important topic of discussion for many years. Recently, increasing attention has been placed upon monitoring the quality of healthcare performance. The use of statistical process control techniques to appropriately monitor the quality of healthcare has garnered increasing interest due to its impact and effectiveness in other fields such as manufacturing and public health surveillance. One characteristic of interest which has received significant attention is that of surgical outcome quality. The timely detection of any deterioration in surgical performance is the key to prompt investigations of possible causes and to avoid adverse consequences. Moreover, it is quite beneficial for healthcare practitioners and organizations to obtain evidence of improved surgical performance.

The cumulative sum (CUSUM) chart is widely used for process monitoring in industrial quality control applications. For monitoring a sequence of independent Bernoulli trials, the Bernoulli CUSUM chart proposed by Reynolds and Stoumbos (1999) is based on the assumption that the Bernoulli proportion is constant when the process is stable. These authors did not allow for varying in-control probability values since industrial items being produced are typically assumed to be relatively homogeneous. The CUSUM chart designed by Reynolds and Stoumbos (1999) is generally not appropriate for monitoring surgical performance when the pre-operative risks vary considerably among patients. Driven by the need for appropriately monitoring surgical performance in the presence of risk variations in the underlying patient populations, the risk-

adjusted Bernoulli CUSUM chart was developed by Steiner *et al*. (2000). With this chart one adjusts for each patient's pre-operative risk of surgical failure using a logistic regression model and then applying the likelihood ratio based scoring method to obtain the monitoring statistics. It has been shown that the risk-adjusted CUSUM chart is suitable for detecting improvement or deterioration in surgical performance when there is a mix of patients with varying pre-operative risks. A general review of risk-adjusted charting was provided by Cook *et al.* (2008). Woodall *et al.* (2015) gave a thorough review on monitoring and improving surgical outcome quality.

A number of practitioners and researchers have applied the risk-adjusted Bernoulli CUSUM method for monitoring clinical outcomes. Sherlaw-Johnson (2005) advised using the signal rule of the risk-adjusted CUSUM in the background with the perhaps more commonly used variable life adjusted display (VLAD). This is the approach advocated by the Clinical Practice Improvement Centre (2008). Sherlaw-Johnson *et al*. (2005) combined the risk-adjusted CUSUM chart and the rocket tail chart based on the VLAD for monitoring by developing a scheme in which CUSUM signals were superimposed onto the rocket tail plots. In another application, Harris *et al*. (2005) applied the risk-adjusted CUSUM method retrospectively to the analysis of their medical center's experience with ruptured abdominal aortic aneurysms (RAAAs) while adjusting for the variability in patients' comorbidities and hemodynamic instability. As another example, Novick *et al*. (2006) compared risk-adjusted and non-risk-adjusted CUSUM analyses of coronary artery bypass surgery outcomes and found that the risk-adjusted CUSUM method was advantageous over non-risk-adjusted methods by not incorrectly signaling a deterioration in performance when preoperative patient risk was high. As a final example, Moore *et al*. (2007) used the risk-adjusted CUSUM method to assess shifts in the performance of an RAAA program over time.

Setting appropriate control limits to get desired in-control average run length (usually denoted as ARL_0) is of importance in the design of any CUSUM procedure. The run length is defined in surgical performance monitoring as the number of Bernoulli trials (i.e., patients) observed until a signal is given by the control chart indicating that a process change has occurred. A relatively large ARL_0 value is desirable when there are no changes in the parameter of interest, whereas a small out-of-control average run length value when the parameter of interest has actually changed indicates good chart performance. Moreover, getting the same or very close ARL_0 values is the prerequisite for comparing the out-of-control performance of competing control charts. In the cardiac surgery example shown in Steiner *et al.* (2000), the control limits for the proposed CUSUM charts were set at a specified level to give a relatively large ARL_0 value given the patient population and the fitted logistic regression model used for risk-adjustment.

However, concerns about the effect of different risk distributions on the performance of risk-adjusted Bernoulli CUSUM charts have been brought up by several researchers. Steiner *et al.* (2001) showed that the in-control average run lengths (ARLs) of risk-adjusted CUSUM charts with the same risk adjustment model and fixed control limits can vary by a factor of 10 for sequences of the highest and lowest risk patients. Thus, they suggested that the control limit of the monitoring procedure be adjusted if the patient mix changes dramatically. Most recently, Tian *et al.* (2015) examined the effect of varying patient population distributions on the in-control performance of the risk-adjusted Bernoulli CUSUM charts. Figure 1.1 shows the five different patient populations examined in this study. Here the patient populations are represented by their Parsonnet score distributions. The simulation results are shown in Table 1.1 and Figure 1.2. It can be clearly seen that the in-control ARLs of the risk-adjusted Bernoulli CUSUM chart

with constant control limits and a given risk-adjustment model can vary by a factor of two for different realistic patient populations.

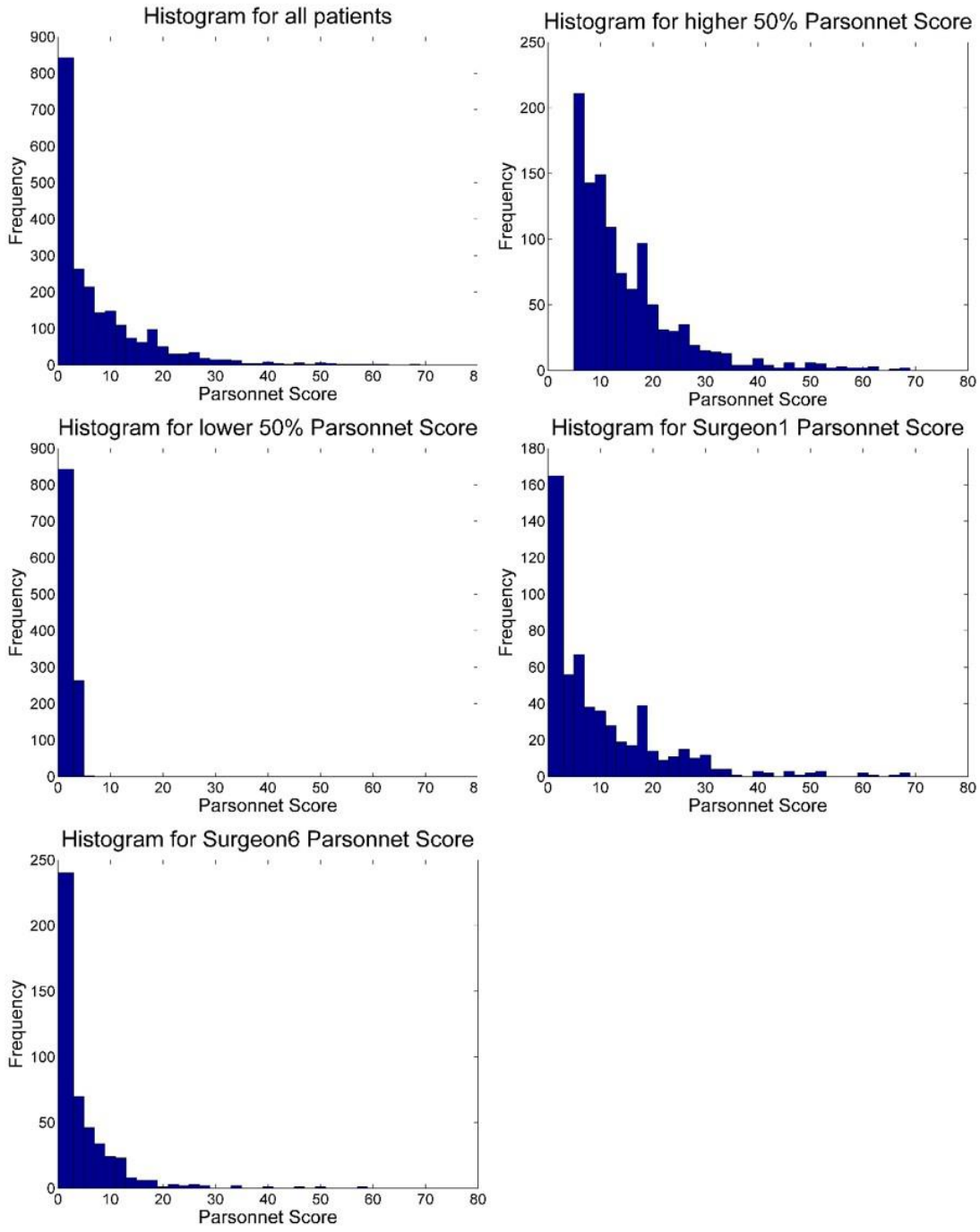


Figure 1.1. Parsonnet score histograms for five different patient populations studied in Tian *et al.* (2015) (used with permission of Tian *et al.*)

Table 1.1. Population means and in-control ARL comparison in Tian *et al.* (2015) (used with permission of Tian *et al.*)

Population	Mean Parsonnet Score	Upper CUSUM ARL_0 (S.E.)	Lower CUSUM ARL_0 (S.E.)
All Phase I scores	8.9026	7,400.1 (73.6)	6,069.3 (59.6)
Lower 50% of scores	2.0541	12,324.0 (120.1)	11,014.0 (105.5)
Higher 50% of scores	16.4813	4,988.6 (50.9)	3,983.7 (39.8)
Surgeon 1 scores	11.2513	6,474.1 (63.9)	5,148.4 (50.2)
Surgeon 6 scores	5.5591	9,657.8 (94.2)	8,207.2 (78.7)

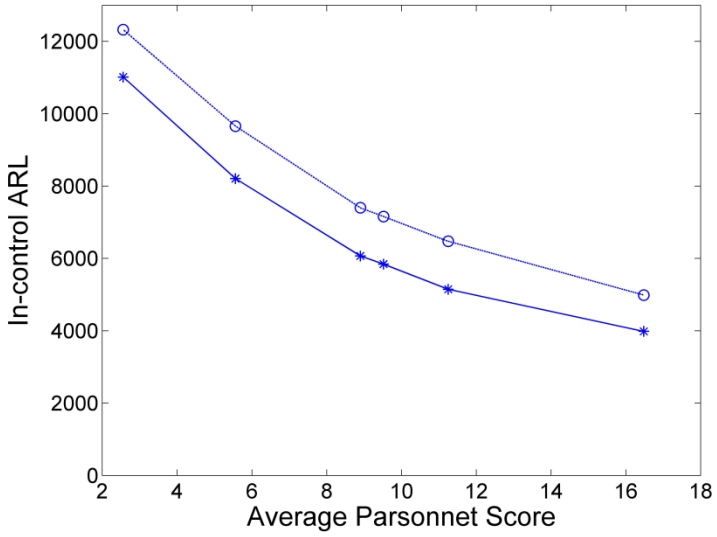


Figure 1.2. The in-control ARLs of upper (circled) and lower (starred) risk-adjusted CUSUM charts given varying risk distributions in Tian *et al.* (2015) (used with permission of Tian *et al.*)

Varying patient populations are very common in applications. The patient mixes for different hospitals and different surgeons can vary considerably and vary over time. Therefore, use of constant control limits for risk-adjusted Bernoulli CUSUM charts leads to quite variable in-control ARL performance. The chart would need to be designed for each surgeon or hospital based on assumptions about the patient population that may be difficult to justify.

To overcome this problem, the method to determine dynamic probability control limits (DPCLs) for the risk-adjusted Bernoulli CUSUM charts is proposed in Chapter 2. The dynamic method of determining probability control limits was proposed by Shen *et al.* (2013) for an application involving the exponentially weighted moving average (EWMA) chart to monitor Poisson count data with time-varying population sizes. The concept in non-dynamic applications was previously used by Margavio *et al.* (1995) and Hawkins (2003). By maintaining the conditional false alarm rates given there is no false for previous observations at a constant level, the risk-adjusted Bernoulli CUSUM chart with DPCLs can consistently deliver a desirable in-control performance with approximately geometrically distributed run lengths for any sequence of patients without regard to the population(s) from which the patients come. Therefore, the use of DPCLs instead of constant control limits for risk-adjusted Bernoulli CUSUM charts is more practical since no assumptions about patients' risk distributions are needed. This study is a result of joint work with Dr. William H. Woodall (Department of Statistics, Virginia Tech).

Chapter 3 is an extended work of Chapter 2 which applies DPCLs to the lower risk-adjusted Bernoulli CUSUM charts to monitor the improvement of surgical outcome quality. The in-control performance of two-sided risk-adjusted Bernoulli CUSUM charts with DPCLs is also examined in Chapter 3. This study is a result of joint work with Dr. William H. Woodall (Department of Statistics, Virginia Tech).

Jones and Steiner (2000) studied the effect of estimation error on the performance of risk-adjusted Bernoulli CUSUM chart with constant control limits systematically and found that the effect could be substantial. Chapter 4 focuses on examining the estimation error on in-control performance of risk-adjusted Bernoulli CUSUM chart with DPCLs. In practice, different Phase I data would provide different parameter estimates and therefore different risk-adjustment models.

The in-control performance of risk-adjusted Bernoulli CUSUM chart with DPCLs would also be affected by this estimation error, but it will not be affected by changes in the patient risk distribution because the control limits are determined depending on the specific observed sequence of patient risk scores. To make a fair comparison, we use the same data and settings as in the study of Jones and Steiner (2000) except for the way to determine the control limits. This study is a result of joint work with Dr. William H. Woodall (Department of Statistics, Virginia Tech).

Tang *et al.* (2015) developed the risk-adjusted CUSUM chart based on multiresponses for monitoring a surgical process with three or more outcomes. Similar to the risk-adjusted Bernoulli CUSUM chart, there is a significant effect of varying risk distributions on the in-control performance of risk-adjusted CUSUM chart for multiresponses when constant control limits are applied. In Chapter 5, the DPCLs approach is adapted to design the risk-adjusted CUSUM chart based on multiresponses and the in-control performance of the charts with DPCLs is examined. This study is a result of joint work with Dr. William H. Woodall (Department of Statistics, Virginia Tech) and Mr. Justin Loda (Department of Statistics, Virginia Tech).

The next chapter, Chapter 2, is the manuscript titled "Dynamic probability control limits for risk-adjusted Bernoulli CUSUM charts" which has been published in *Statistics in Medicine* (2015). This is followed by Chapter 3, the manuscript titled "Dynamic probability control limits for lower and two-sided risk-adjusted Bernoulli CUSUM charts" which has been submitted to *Journal of Quality Technology*. Next, Chapter 4 is the manuscript titled "The effect of estimation error on in-control performance of risk-adjusted Bernoulli CUSUM chart with dynamic probability control limits" which has been submitted to *The International Journal of Biostatistics* and Chapter 5 is the manuscript titled "Dynamic probability control limits for risk-adjusted

CUSUM charts based on multiresponses" which has been submitted to *Journal of the American Statistical Association*. Finally, the conclusions and some future research ideas are discussed in Chapter 6.

Chapter 2

Dynamic probability control limits for risk-adjusted Bernoulli CUSUM charts

ABSTRACT

The risk-adjusted Bernoulli cumulative sum (CUSUM) chart developed by Steiner et al. (2000) is an increasingly popular tool for monitoring clinical and surgical performance. In practice, however, use of a fixed control limit for the chart leads to quite variable in-control average run length (ARL) performance for patient populations with different risk score distributions. To overcome this problem, we determine simulation-based dynamic probability control limits (DPCLs) patient-by-patient for the risk-adjusted Bernoulli CUSUM charts. By maintaining the probability of a false alarm at a constant level conditional on no false alarm for previous observations, our risk-adjusted CUSUM charts with DPCLs have consistent in-control performance at the desired level with approximately geometrically distributed run lengths. Our simulation results demonstrate that our method does not rely on any information or assumptions about the patients' risk distributions. The use of DPCLs for risk-adjusted Bernoulli CUSUM charts allows each chart to be designed for the corresponding particular sequence of patients for a surgeon or hospital.

Keywords: average run length (ARL); false alarm rate; run length distribution; statistical process control; surgical performance.

1. Introduction

Appropriately monitoring the quality of healthcare performance has become increasingly important. In particular, monitoring the quality of surgical outcome performance has been receiving increasing interest [1, 2]. It is important to detect any deterioration in surgical performance in a timely manner to avoid adverse consequences. Also, it is quite beneficial to obtain evidence of improved surgical performance.

Generally, there is considerable variability in prior risks of different patient populations. Thus, the standard Bernoulli cumulative sum (CUSUM) chart which does not allow for heterogeneity in patients cannot be applied effectively to monitor surgical performance. Steiner *et al.* [3] developed the risk-adjusted Bernoulli CUSUM chart which involves adjusting for each patient's pre-operative risk of surgical failure using a logistic regression model and then applying a likelihood-ratio based scoring method to obtain the monitoring statistics. It has been shown to be useful for detecting surgical deterioration or improvement in settings where there is a mix of patients with varying pre-operative risks. The risk-adjusted Bernoulli CUSUM chart has been recommended by several researchers [2, 4-6] due to its advantages of design, performance and effectiveness in reducing the effect of patients' prior risks. In practice, a number of practitioners have applied this chart for various applications. Some of them use the risk-adjusted CUSUM charts directly to assess the performance of certain kinds of operations [7-13] or to monitor clinical performance of healthcare organizations [14], while some run a risk-adjusted CUSUM chart with the more easily interpretable variable life adjusted display (VLAD) chart for monitoring clinical outcomes [15-19].

However, concerns about the effect of different risk distributions on the performance of risk-adjusted Bernoulli CUSUM charts have been brought up by several researchers. Steiner *et al.* [4] showed that the in-control average run lengths (ARLs) of risk-adjusted CUSUM charts with the same risk adjustment model and fixed control limits can vary by a factor of 10 for sequences of the highest and lowest risk patients. Thus, they suggested that the control limit of the monitoring procedure be adjusted if the patient mix changes dramatically. The ARL is the expected number of patients until a control chart signal is given. Rogers *et al.* [1] stated that a design based on prior data might not yield an in-control ARL accurately if the risk profile changes over time. In addition, Loke and Gan [20] compared various theoretical risk distributions and found that the in-control ARLs of risk-adjusted CUSUM charts were clearly affected by changes in the risk distribution. In fact, they recommended a method based on the beta distribution for monitoring the risk distribution. More recently, Tian *et al.* [21] showed that the in-control ARLs of risk-adjusted CUSUM charts with fixed control limits and a given risk-adjustment model can vary by a factor of two for different realistic patient populations.

Varying patient populations are very common in applications. The patient mixes for different hospitals and different surgeons can vary considerably and vary over time. Therefore, use of fixed control limits for risk-adjusted Bernoulli CUSUM charts leads to quite variable in-control ARL performance. The chart would need to be designed for each surgeon or hospital based on assumptions about the patient population that may be difficult to justify.

To address this problem, we propose to apply dynamic probability control limits (DPCLs) to the risk-adjusted Bernoulli CUSUM chart for monitoring the surgical performance for specific sequences of patients. The dynamic method of determining probability control limits was proposed by Shen *et al.* [22] for an application involving the exponentially weighted moving

average (EWMA) chart to monitor Poisson count data with time-varying population sizes. The concept in non-dynamic applications was previously used by Margavio *et al.* [23] and Hawkins [24]. The idea is to maintain the conditional probability of a false alarm given there is no false alarm for previous observations at a constant value. By applying the DPCLs to the risk-adjusted Bernoulli CUSUM chart, one can obtain an in-control run length distribution which is approximately a geometric distribution with a desired in-control ARL for any sequence of patients without regard to the population(s) from which the patients come. Thus, our method overcomes the primary shortcoming of the risk-adjusted CUSUM chart, therefore making it more practical for monitoring surgical performance since no assumptions about patients' risk distributions are needed. It is important to note that although we control the conditional false alarm rate to be constant, our method can be used to design the chart with any specified sequence of conditional false alarm rates. We note that the approach of Gombay *et al*. [25] is different from ours in the sense that they control the overall probability of a false alarm for a specified number of patients. We instead assume an ongoing monitoring scenario.

The remainder of this paper is organized as follows. The risk-adjusted Bernoulli CUSUM chart and the proposed simulation-based procedure for determining the DPCLs are introduced in Section 2. Next, the simulation using the data set examined by Steiner *et al.* [3] to validate our method is explained in Section 3. This is followed by results demonstrating the desirable in-control run length distributions obtained by applying the proposed DPCLs and a comparison of out-of-control performance with the use of fixed control limits in Section 4. Finally, we give our conclusions and make some related remarks in Section 5.

2. Method

2.1. Risk-adjusted Bernoulli CUSUM chart

The one-sided tabular CUSUM statistics can be written as follows,

$$C_t = \max\left(0, C_{t-1} + W_t\right),\ t = 1, 2, 3, \ldots \tag{2.1}$$

where $C_0 = 0$ and W_t represents the weight assigned to the result for the t^{th} individual. To monitor a specified change of the parameter of interest θ, say from θ_0 to θ_a, the optimal choice for W_t is the log-likelihood ratio $W_t = \ln\left(f\left(y_t;\theta_a\right)/f\left(y_t;\theta_0\right)\right)$, where y_t is the outcome for the t^{th} individual. The two-sided CUSUM chart is obtained by running two one-sided charts simultaneously, one with $\theta_a < \theta_0$ and one with $\theta_a > \theta_0$. The signs of the statistics can be changed for one of the sets of cumulative sum statistics for ease of plotting the charts one above the other.

Driven by the need to take into account the pre-operative risks of an adverse event of interest which vary considerably among different patients in practice, Steiner *et al.* [3] developed a monitoring approach in which one can adjust for each patient's prior risk of surgical failure. It is thus referred to as the risk-adjusted Bernoulli CUSUM chart. By denoting the adjusted surgical failure rate for patient *t* by $p_t\left(\theta\right)$, we obtain $f\left(y_t \mid \theta\right) = p_t\left(\theta\right)^{y_t}\left[1 - p_t\left(\theta\right)\right]^{1-y_t}$, where y_t = 1 if patient *t* experiences the adverse event of interest, such as death or a surgical site infection within a specified time period following surgery, and y_t = 0 otherwise.

The surgical failure rate for each patient is determined by assessing the pre-operative risk of each patient by applying a method such as a logistic regression model based on Parsonnet

scores [26]. Since $p_t(\theta)$ usually varies considerably from patient to patient, monitoring for any change in $p_t(\theta)$ is not useful. We let *R* denote the odds ratio corresponding to failure. Then for patient *t*, $Rp_t : (1-p_t)$ are the odds of failure and the corresponding probability of failure is $Rp_t/(1-p_t+Rp_t)$. The risk-adjusted CUSUM chart is designed to monitor for an odds ratio change from R_0 to R_a, where R_0 is usually set to 1 to reflect current expected surgical performance, and set to a value $R_a > R_0$ for detecting performance deterioration, and to a value $R_a < R_0$ to detect process improvement. Thus, the risk-adjusted CUSUM chart weights can be calculated as

$$W_t = \begin{cases} \log\left(\dfrac{1-p_t+R_0p_t}{1-p_t+R_ap_t}\right) & \text{if } y_t = 0, \\ \log\left[\dfrac{(1-p_t+R_0p_t)R_a}{(1-p_t+R_ap_t)R_0}\right] & \text{if } y_t = 1. \end{cases} \quad . \qquad (2.2)$$

We let C_t^+, $t = 1, 2, \ldots$, represent the CUSUM statistics designed to detect an increase in the odds ratio and C_t^-, $t = 1, 2, \ldots$, the CUSUM statistics designed to detect a decrease in the odds ratio. Following the convention we change the signs of the statistics for the lower CUSUM chart and do not allow the values to exceed zero. The chart signals when $C_t^+ > h^+$ or $C_t^- < h^-$ which indicates that there has been either deterioration or improvement in the surgical performance, respectively. The control limits h^+ and h^- are set to yield a suitably large in-control ARL when there are no changes in the odds ratio of failure R_0.

However, as discussed in Section 1, the in-control ARLs have been found to vary significantly as the patient risk distributions change if we apply fixed control limits to the risk-

adjusted Bernoulli CUSUM chart. Moreover, accurate information about the patient risk distributions, which determines the design of the chart, is usually unavailable in practice. In addition, the risk distribution may change over time, as it would tend to do for a beginning surgeon who gains more and more experience. More experienced surgeons tend to operate on sicker patients than new surgeons. To solve this design problem, we propose the use of dynamic probability control limits for risk-adjusted CUSUM charts in the next section.

2.2. Dynamic probability control limits (DPCLs)

The method of determining dynamic probability control limits was developed by Shen *et al.* [22] for the EWMA chart for monitoring Poisson count data with time-varying sample sizes. The main idea is to keep the probability of obtaining a false alarm constant from sample to sample conditional on no false alarm for the previous observations. Specifically, for a one-sided risk-adjusted Bernoulli CUSUM chart, the DPCLs $\boldsymbol{h}(\alpha)=\left(h_1(\alpha),\ h_2(\alpha),\ ...,\ h_k(\alpha),\ ...\right)$ satisfy to the extent possible the following equations:

$$\begin{cases} \Pr\left(C_1^+ > h_1(\alpha) \middle| S_1\right) = \alpha, \\ \Pr\left(C_t^+ > h_t(\alpha) \middle| C_k^+ \le h_k(\alpha),\ k=1,\ ...,\ t-1,\ S_t\right) = \alpha, \text{ for } t=2,\ 3,\ \ldots, \end{cases} \tag{2.3}$$

where $S_1,\ S_2,\ \ldots$ is the sequence of observed risk scores, such as Parsonnet scores, and α is the predetermined conditional false alarm rate. For each patient, the observed risk-adjusted CUSUM statistic $C_t^+ = \max\left(0, C_{t-1}^+ + W_t\right)$ is calculated, where W_t is computed using equation (2.2) based on the observed y_t and the value of p_t. The monitoring process signals when $C_t^+ > h_t(\alpha)$. From equation (2.3), it follows that the in-control run length is approximately the geometric

distribution with parameter $p = \alpha$. Thus, we found the in-control ARL with use of the DPCLs $\boldsymbol{h}(\alpha) = (h_1(\alpha),\ h_2(\alpha),\ ...,\ h_k(\alpha),\ ...)$ to be close to $1/\alpha$.

The computational procedure to obtain the DPCLs is outlined below. For simplicity, we only explain this procedure for the one-sided risk-adjusted Bernoulli CUSUM chart for detecting process deterioration, *i.e.* $R_a > R_0$. This computational algorithm can be applied similarly to the other side of a two-sided risk-adjusted Bernoulli CUSUM chart for also detecting shifts to a value $R_a < R_0$.

In general terms, the iterative algorithm works as follows: we simulate a large number of CUSUM statistics for each patient. For a given patient, we randomly sample a large number of values of the immediately previous CUSUM statistics from our simulation which did not result in an out-of-control signal. We combine these values with randomly generated Bernoulli random variables with the parameter value based on the current patient's risk score. We then choose an upper percentile of these updated CUSUM values as the control limit for the current patient.

At the beginning, consider the first patient with some risk score S_1. Under expected performance, the estimated rate p_1 can be calculated from the risk-adjustment model. Next, we generate N random variables $Y_{1,i}$ which are independent and identically distributed (i.i.d.) Bernoulli random variables with probability p_1, $i = 1, 2, \ldots, N$, where N is a sufficiently large number. Then we calculate the risk-adjusted CUSUM statistics $C_{1,1}$, $C_{1,2}$, $C_{1,3}$, …, $C_{1,N}$ corresponding to the $Y_{1,i}$'s, using the equations (2.1) and (2.2), sort the N CUSUM statistics in ascending order $C_{1,(1)}$, $C_{1,(2)}$, $C_{1,(3)}$, …, $C_{1,(N)}$ and take the $N' = \lceil N \cdot (1-\alpha) \rceil$ largest CUSUM statistic $C_{1,(N')}$ as the approximated DPCL $h_1(\alpha)$. Note that the CUSUM statistic can only take

two values at t = 1 (i.e. the first patient) due to the binary property of Bernoulli variables. Note also that for the first few patients there may not be a control limit such that the approximated conditional false alarm rate is α or below. In these cases, there is no control limit used and no possibility of the chart signaling.

Then, for patient 2 with risk score S_2, we generate N i.i.d. Bernoulli random values $Y_{2,i}$ with probability p_2, where p_2 is calculated from our logistic regression model. Again, N corresponding risk-adjusted CUSUM statistics $C_{2,i}$, i = 1, 2, …, N are computed from equations (2.1) and (2.2) with randomly chosen values from a vector of $C_{1,i}$ values which are less than or equal to $h_1(\alpha)$. Then we sort the N $C_{2,i}$ values in ascending order and take the upper α percentile $C_{2,(N')}$ as the DPCL $h_2(\alpha)$. For the next patient, we keep the CUSUM statistics which are less than or equal to $h_2(\alpha)$ and repeat the same procedure as before to obtain $h_3(\alpha)$. The process is continued iteratively to obtain other control limits. The algorithmic form of the simulation procedure to obtain the control limit for patient t (t = 1, 2, 3, …), can be summarized as follows:

1) Generate N Bernoulli random variables $Y_{t,i}$ (i = 1, 2, …, N) with in-control failure rate p_t obtained from risk-adjustment model and N CUSUM statistics $C_{t,i} = \max\left(0, C_{t-1,j} + W_{t,i}\right)$ (j = 1, 2, …, N) using the equations (2.1) and (2.2) accordingly, where $C_{t-1,j}$ is randomly selected from a vector of CUSUM values $\boldsymbol{C}_{\boldsymbol{t\text{-}1}}{}' = \left\{C_{t-1,i} \text{ such that } C_{t-1,i} \le h_{t-1}(\alpha)\right\}$.
2) Sort the N CUSUM statistics in ascending order $C_{t,(1)}$, $C_{t,(2)}$, $C_{t,(3)}$, …, $C_{t,(N)}$ and take the $N' = \left\lceil N \cdot (1-\alpha) \right\rceil$ largest CUSUM statistic $C_{t,(N')}$ as the approximated DPCL $h_t(\alpha)$.

3) Calculate $C_t^+ = \max\left(0, C_{t-1}^+ + W_t\right)$ based on the surgical outcome and risk score of patient *t*.

If $C_t^+ > h_t(\alpha)$, an out-of-control signal is issued. Otherwise, go back to step 1.

Due to the discreteness of Bernoulli random variables, for each patient t = 1, 2, … , $\Pr\left\{C_{t,i} > h_t(\alpha)\right\}$ (denoted by α_t) is always controlled to be less than or equal to α and $\Pr\left\{C_{t,i} \geq h_t(\alpha)\right\}$ (denoted by α_t') to be always greater than or equal to α. As the control limit determination procedure progresses, an increasing number of different CUSUM statistic values will be generated. After a relatively small number of patients, we will observe more $\Pr\left\{C_{k,i} > h_k(\alpha)\right\}$ values close to α. As shown in Section 4, the conditional false alarm rate, however, does not converge to α as the number of patients increases.

3. Simulation settings

The previous study of Tian *et al.* [21] found that the in-control ARL of the Bernoulli risk-adjusted CUSUM chart with fixed control limits varies considerably for different patient populations. Specifically, the in-control ARLs decrease as the mean risk score of the patient population increases. In our simulation work, we examined the in-control performance of the Bernoulli risk-adjusted CUSUM chart with our proposed DPCLs for specific sequences of patients from different patient populations.

We used the same data set of patients from a seven-year study used by Steiner *et al.* [3]. The 2,218 patients from the first two years were treated as in-control group and used to fit the following logistic regression risk model:

$$\text{logit}\left(p_t\right) = -3.68 + 0.077 X_t, \tag{2.4}$$

where X_t is the Parsonnet score of patient t and p_t is the probability of death within thirty days following surgery. Since the Parsonnet score is the only explanatory variable in this logistic regression model to determine the probability of death, we can use the different Parsonnet score distributions to represent the different patient populations. We also use the same criteria as Tian *et al.* [21] to differentiate the Parsonnet score distributions. We randomly chose several sequences of 20,000 Parsonnet scores with replacement from each of the following five different Phase I risk distributions:

1) All: 2,218 scores for all patients (Mean = 8.9026),
2) High Risk: the highest 50% of the scores (Mean = 16.4813),
3) Low Risk: the lowest 50% of the scores (Mean = 2.0541),

4) Surgeon 1: 565 scores for all of surgeon 1's patients (Mean = 11.2513),

5) Surgeon 6: 474 scores for all of surgeon 6's patients (Mean = 5.5591).

Then for each sequence of scores, we calculated the DPCLs applying the algorithm explained in Section 2.2. Here, the conditional false alarm rate α was chosen and N, the number of CUSUM statistics simulated, was set to be a suitably large number. The risk-adjusted CUSUM statistics were obtained using equations (2.1) and (2.2) with $R_0 = 1$ and $R_a = 2$. We also recorded $\Pr\{C_{t,i} > h_t(\alpha)\}$ and $\Pr\{C_{t,i} \geq h_t(\alpha)\}$ for $t = 1, 2, \ldots$ as a check for accuracy of the procedure. After we obtained a sequence of DPCLs for a specific sequence of scores selected from a particular risk distribution, 100,000 control charts were simulated to estimate the in-control ARL (ARL_0). Also, the 10th, 25th, 50th, 75th, 90th percentiles of the run length ($Q_{0.10}$, $Q_{0.25}$, $Q_{0.50}$, $Q_{0.75}$, $Q_{0.90}$) and standard deviation of the run length (SDRL) were estimated in order to compare with those of the geometric distribution.

4. Results

4.1. Estimated in-control performance

We first set the conditional false alarm rate α to be 0.005. For each of the five risk distributions described in the previous section, ten different sequences of 20,000 scores were randomly chosen and used to construct the control chart. Here we used N = 100,000. The simulation results are listed in Table 2.1 for the first five of the sequences along with the geometric distribution with $p = \alpha = 0.005$ for comparison. Theoretically, the in-control run length of the risk-adjusted Bernoulli CUSUM chart with our proposed DPCLs for any sequence of risk scores should follow an approximate geometric distribution. We note from Table 2.1, that the estimated ARL_0, SDRL and $Q_{0.10}$, $Q_{0.25}$, $Q_{0.50}$, $Q_{0.75}$, $Q_{0.90}$ values for each sequence examined are close to the corresponding theoretical values of a geometric distribution with p = 0.005. The SDRL value is approximately equal to the ARL_0 value for each sequence, as it would be for the geometric distribution. However, we do observe slight deviations of the run length distributions from the theoretical geometric distribution, which is due to the discrete nature of the Bernoulli random variables. The average value of $\Pr\{C_{t,i} > h_t(0.005)\}$ (denoted by $\overline{\alpha_t}$) and the average value of $\Pr\{C_{t,i} \geq h_t(0.005)\}$ (denoted by $\overline{\alpha_t'}$) are also reported for each sequence of scores. We can see that the $\overline{\alpha_t}$ values vary from 0.0047581 to 0.0048162 and the $\overline{\alpha_t'}$ values vary from 0.0052268 to 0.0053224. Thus the averages are close to the desired conditional false alarm rate of α = 0.005.

Table 2.1. Estimated in-control performance of risk-adjusted Bernoulli CUSUM charts with DPCLs (α = 0.005)

Risk distribution	Sequence index	$\widehat{\text{ARL}_0}$(S.E.)	$\widehat{\text{SDRL}}$	$Q_{0.10}$	$Q_{0.25}$	$Q_{0.50}$	$Q_{0.75}$	$Q_{0.90}$	$\overline{\alpha_t}$	$\overline{\alpha_t'}$
All	1	212.7 (0.66)	209.2	27	65	148	291	485	4.7602E-3	5.3073E-3
	2	218.0 (0.67)	211.2	28	67	154	299	493	4.7581E-3	5.3171E-3
	3	215.3 (0.65)	206.7	28	67	152	299	486	4.7601E-3	5.3079E-3
	4	217.0 (0.67)	211.6	27	66	153	300	492	4.7656E-3	5.3064E-3
	5	219.5 (0.67)	211.5	31	69	155	302	495	4.7624E-3	5.3054E-3
High Risk	1	212.6 (0.66)	207.1	29	67	149	291	481	4.8090E-3	5.2320E-3
	2	213.8 (0.66)	207.6	28	66	151	293	484	4.8095E-3	5.2357E-3
	3	214.9 (0.65)	205.5	29	68	153	296	483	4.8084E-3	5.2349E-3
	4	212.2 (0.66)	208.3	25	64	148	293	486	4.8101E-3	5.2402E-3
	5	211.7 (0.65)	206.7	27	64	147	294	482	4.8104E-3	5.2347E-3
Low Risk	1	213.7 (0.66)	207.5	27	66	150	295	484	4.8119E-3	5.2284E-3
	2	211.9 (0.66)	207.5	27	64	148	291	482	4.8134E-3	5.2268E-3
	3	212.9 (0.66)	209.7	27	65	147	294	487	4.8130E-3	5.2302E-3
	4	212.6 (0.66)	208.2	26	65	148	293	484	4.8152E-3	5.2302E-3
	5	213.9 (0.66)	208.8	26	64	151	295	486	4.8139E-3	5.2310E-3
Surgeon 1	1	212.9 (0.66)	208.2	27	65	150	292	483	4.7685E-3	5.2837E-3
	2	215.8 (0.66)	209.6	28	67	154	295	490	4.7700E-3	5.2845E-3
	3	218.3 (0.66)	209.8	30	68	155	299	491	4.7664E-3	5.2876E-3
	4	215.2 (0.66)	208.3	28	68	151	295	484	4.7673E-3	5.3120E-3
	5	217.7 (0.66)	209.4	29	69	153	300	493	4.7670E-3	5.2904E-3
Surgeon 6	1	214.0 (0.66)	207.5	28	66	151	294	484	4.7946E-3	5.2701E-3
	2	212.9 (0.66)	208.5	26	64	149	293	487	4.7884E-3	5.2906E-3
	3	215.2 (0.66)	208.7	28	67	151	297	488	4.7937E-3	5.2817E-3
	4	213.6 (0.66)	207.6	27	65	151	297	483	4.7928E-3	5.2808E-3
	5	215.5 (0.66)	208.9	27	67	152	298	490	4.7952E-3	5.2788E-3
Geometric	/	200 (/)	200	21	57	138	276	459	/	/

Secondly, we considered a lower conditional false alarm rate of α = 0.001. We again randomly chose ten different sequences of 20,000 scores from each of the five risk distributions and conducted the same simulation procedures as with α = 0.005 and N = 100,000. The results are summarized in Table 2.2 for the first five of the sequences and compared with the geometric distribution with $p = \alpha$ = 0.001. The estimated ARL_0, SDRL and $Q_{0.10}$, $Q_{0.25}$, $Q_{0.50}$, $Q_{0.75}$, $Q_{0.90}$ values for each sequence examined are even closer to the theoretical reference values of the

geometric distribution with p = 0.001 with respect to relative error. The SDRL and ARL_0 values are very close to each other for each sequence of scores. The average $\Pr\{C_{t,i} > h_t(0.001)\}$ values ($\overline{\alpha_t}$) vary from 0.00098175 to 0.00099264 and the average $\Pr\{C_{t,i} \geq h_t(0.001)\}$ values ($\overline{\alpha_t'}$) vary from 0.0010195 to 0.0010491. Both averages are very close to the desired conditional false alarm rate of 0.001.

Table 2.2. Estimated in-control performance of risk-adjusted Bernoulli CUSUM charts with DPCLs (α = 0.001)

Risk distribution	Sequence index	$\widehat{ARL_0}$(S.E.)	$\widehat{SDRL}$	$Q_{0.10}$	$Q_{0.25}$	$Q_{0.50}$	$Q_{0.75}$	$Q_{0.90}$	$\overline{\alpha_t}$	$\overline{\alpha_t'}$
All	1	1027.6 (3.26)	1029.6	109	291	714	1432	2363	9.8322E-4	1.0303E-3
	2	1003.5 (3.15)	997.0	111	291	697	1384	2310	9.8314E-4	1.0331E-3
	3	1023.0 (3.18)	1005.4	113	298	721	1424	2340	9.8317E-4	1.0325E-3
	4	1016.7 (3.22)	1019.6	111	292	708	1406	2350	9.8292E-4	1.0317E-3
	5	1020.4 (3.20)	1012.0	114	303	715	1403	2337	9.8265E-4	1.0313E-3
High Risk	1	1015.2 (3.19)	1009.4	118	296	708	1400	2329	9.8222E-4	1.0346E-3
	2	1017.7 (3.18)	1006.6	118	302	711	1404	2330	9.8265E-4	1.0310E-3
	3	1012.7 (3.18)	1005.5	116	300	707	1395	2324	9.8245E-4	1.0491E-3
	4	1017.0 (3.17)	1000.9	111	304	717	1414	2322	9.8213E-4	1.0324E-3
	5	1023.6 (3.21)	1015.4	113	300	721	1411	2338	9.8276E-4	1.0403E-3
Low Risk	1	1006.4 (3.19)	1008.1	111	289	696	1388	2314	9.9257E-4	1.0196E-3
	2	992.5 (3.14)	992.4	107	287	689	1370	2287	9.9256E-4	1.0207E-3
	3	1013.1 (3.20)	1011.2	112	288	702	1403	2338	9.9258E-4	1.0195E-3
	4	1021.8 (3.19)	1007.9	114	299	712	1419	2330	9.9264E-4	1.0208E-3
	5	1007.8 (3.16)	999.4	108	295	707	1397	2313	9.9243E-4	1.0209E-3
Surgeon 1	1	1015.4 (3.18)	1006.0	112	296	704	1409	2330	9.8256E-4	1.0320E-3
	2	1027.0 (3.21)	1015.1	119	303	716	1427	2355	9.8211E-4	1.0345E-3
	3	1016.7 (3.20)	1011.8	117	298	703	1412	2324	9.8175E-4	1.0316E-3
	4	1032.7 (3.22)	1018.8	120	301	722	1427	2359	9.8180E-4	1.0329E-3
	5	1027.5 (3.21)	1015.6	121	307	722	1411	2343	9.8198E-4	1.0345E-3
Surgeon 6	1	1018.0 (3.22)	1017.2	112	293	708	1400	2342	9.8690E-4	1.0253E-3
	2	1028.8 (3.22)	1018.6	117	302	719	1424	2353	9.8682E-4	1.0266E-3
	3	1018.3 (3.18)	1004.5	114	299	714	1405	2343	9.8696E-4	1.0263E-3
	4	1009.6 (3.18)	1004.5	111	297	699	1387	2329	9.8693E-4	1.0277E-3
	5	1008.6 (3.17)	1003.4	116	298	702	1391	2321	9.8668E-4	1.0259E-3
Geometric	/	1000 (/)	1000	105	287	692	1385	2301	/	/

We also decreased both of the conditional false alarm rates by a factor of 10 and examined the in-control performance with the proposed DPCLs. For each of $\alpha = 0.0005$ and $\alpha = 0.0001$, we randomly chose two sequences of 200,000 scores from each of the 5 risk distributions. Here, we set N = 1,000,000 and implemented the same simulation procedures as before for each sequence. The in-control performance is summarized in Table 2.3 and Table 2.4, respectively. The results show that our method works quite well for even lower desired conditional false alarm rates.

Table 2.3. Estimated in-control performance of risk-adjusted Bernoulli CUSUM charts with DPCLs ($\alpha = 0.0005$)

Risk distribution	Sequence index	$\widehat{\text{ARL}}_0$(S.E.)	$\widehat{\text{SDRL}}$	$Q_{0.10}$	$Q_{0.25}$	$Q_{0.50}$	$Q_{0.75}$	$Q_{0.90}$	$\overline{\alpha_t}$	$\overline{\alpha_t'}$
All	1	2018.0(6.33)	2003.2	221	593	1398	2797	4622	4.9698E-4	5.0440E-4
	2	2011.3(6.33)	2002.8	220	584	1394	2788	4617	4.9701E-4	5.0480E-4
High Risk	1	2020.7(6.33)	2001.1	228	593	1403	2795	4633	4.9573E-4	5.0599E-4
	2	2017.4(6.35)	2008.7	224	591	1400	2798	4629	4.9573E-4	5.0596E-4
Low Risk	1	1999.2(6.32)	1998.3	216	579	1384	2761	4603	4.9918E-4	5.0200E-4
	2	2009.0(6.34)	2005.8	220	582	1402	2766	4599	4.9918E-4	5.0197E-4
Surgeon 1	1	2026.3(6.31)	1994.6	225	593	1424	2809	4643	4.9642E-4	5.0482E-4
	2	2036.1(6.40)	2023.9	222	594	1411	2820	4682	4.9644E-4	5.0482E-4
Surgeon 6	1	2006.1(6.32)	1997.4	222	587	1391	2777	4614	4.9810E-4	5.0305E-4
	2	2009.1(6.34)	2005.8	221	586	1393	2776	4603	4.9814E-4	5.0308E-4
Geometric	/	2000 (/)	2000	210	575	1385	2771	4604	/	/

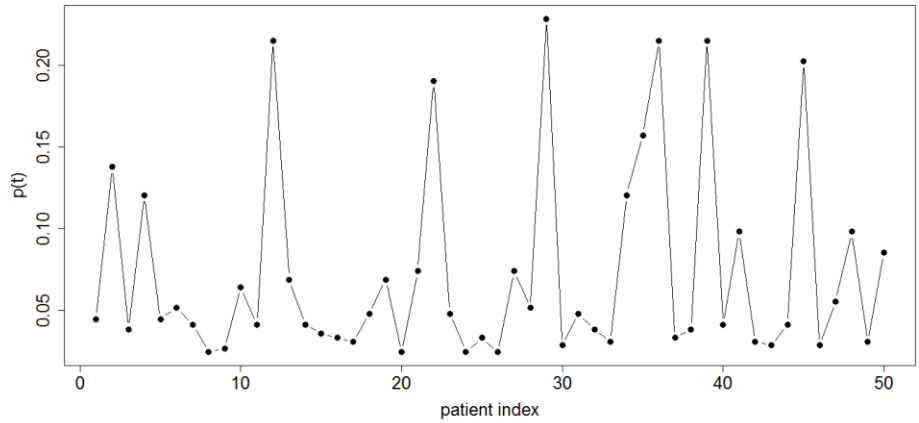
Table 2.4. Estimated in-control performance of risk-adjusted Bernoulli CUSUM charts with DPCLs (α = 0.0001)

Risk distribution	Sequence index	$\widehat{ARL_0}$(S.E.)	$\widehat{SDRL}$	$Q_{0.10}$	$Q_{0.25}$	$Q_{0.50}$	$Q_{0.75}$	$Q_{0.90}$	$\overline{\alpha_t}$	$\overline{\alpha_t'}$
All	1	10027.2(31.66)	10011.6	1044	2869	6985	13902	23029	9.9069E-5	1.0214E-4
	2	10114.1(31.97)	10110.0	1041	2896	7006	14059	23320	9.9067E-5	1.0298E-4
High Risk	1	10107.4(32.06)	10139.5	1059	2908	6977	14024	23195	9.9120E-5	1.0260E-4
	2	10076.9(31.81)	10057.8	1067	2879	6973	13958	23300	9.9114E-5	1.0255E-4
Low Risk	1	9973.5(31.62)	9999.2	1046	2861	6929	13782	23004	9.9618E-5	1.0154E-4
	2	9996.2(31.64)	10006.2	1068	2896	6936	13831	23076	9.9620E-5	1.0157E-4
Surgeon 1	1	10088.5(31.84)	10069.1	1063	2894	6989	14012	23240	9.8961E-5	1.0222E-4
	2	10024.4(31.56)	9980.5	1079	2888	6997	13933	22960	9.8962E-5	1.0229E-4
Surgeon 6	1	10049.3(31.65)	10007.1	1082	2916	6961	13945	23069	9.9329E-5	1.0186E-4
	2	9981.0(31.69)	10021.5	1049	2851	6920	13801	23039	9.9327E-5	1.0193E-4
Geometric	/	10000 (/)	10000	1053	2876	6931	13862	23024	/	/

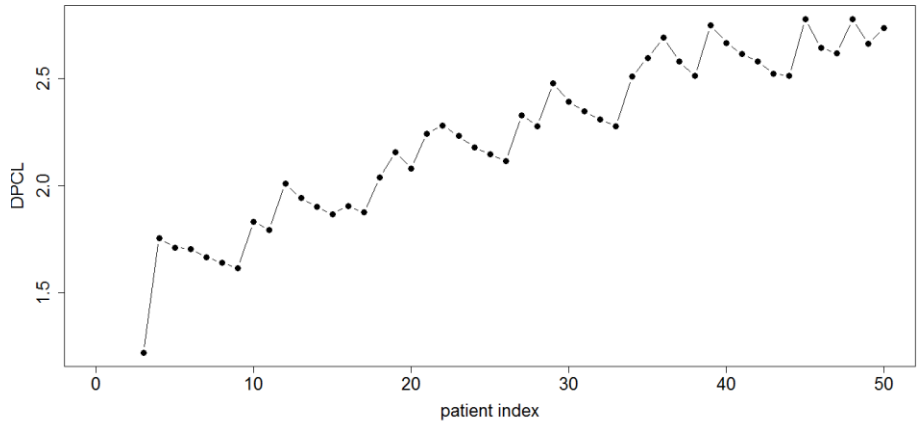
Gombay *et al*. [25] designed some sequential curtailed and risk-adjusted charts to control the overall false alarm rate over a specific horizon of patients. Our method could be used to achieve the same goal since the in-control run lengths resulted from DPCLs approach are approximately geometrically distributed for any sequence of patients. Therefore, the overall false alarm rate $\alpha_{overall}$ for a sequence of T patients can be approximated by $1-(1-\alpha)^T$. Then for a specified $\alpha_{overall}$, one could set the conditional false alarm rate to be $1-(1-\alpha_{overall})^{1/T}$. For instance, if T = 100 and we desired $\alpha_{overall}$ = 0.01, the conditional false alarm rate could be set as $1-(1-0.01)^{1/100} \approx 0.0001$. Our simulation results showed that the average overall false alarm rate $\alpha_{overall}$ for the first T = 100 patients of all 10 sequences with α = 0.0001 was 0.009026 (S.E. = $1.644E^{-4}$) which is close to 0.01.

To demonstrate the application of the proposed DPCLs, we plotted some results of the first sequence from the risk distribution "Surgeon 1" with the conditional false alarm rate set as 0.001 in Figures 2.1 and 2.2.

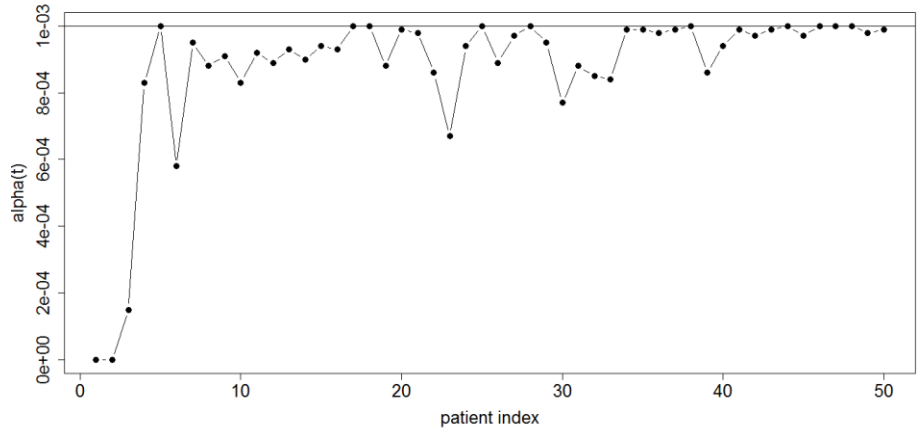
Figure 2.1 gives the results for the first 50 patients. Figure 2.1(a) shows that the surgical failure rates corresponding to the 50 patients' risk scores vary considerably. Figure 2.1(b) shows the DPCLs obtained using our proposed procedure. There are no control limits for the first two observations. The reason is that the approximated conditional false alarm rate cannot be set to be less than or equal to α. Thus, there cannot be any signals at the first two observations for this particular sequence. The solid circles indicate the DPCLs for each of the patients. We see that the probability control limits vary dynamically from patient to patient. Figure 2.1(c) illustrates the conditional false alarm rates $\alpha_t = \Pr\{C_{t,i} > h_t(0.001)\}$ which are always controlled to be less than or equal to α = 0.001. Similar results are shown in Figure 2.2 which provides the results for the first 1,000 patients in the same sequence.



(a) Estimated surgical failure rate p_t

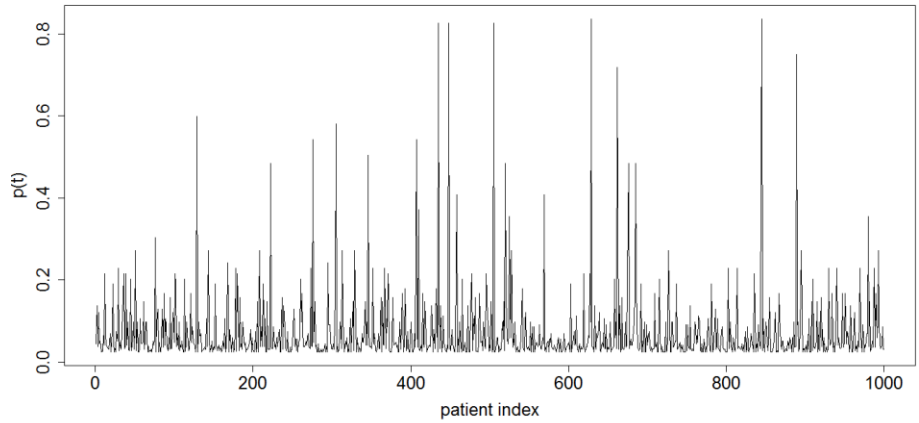


(b) DPCLs

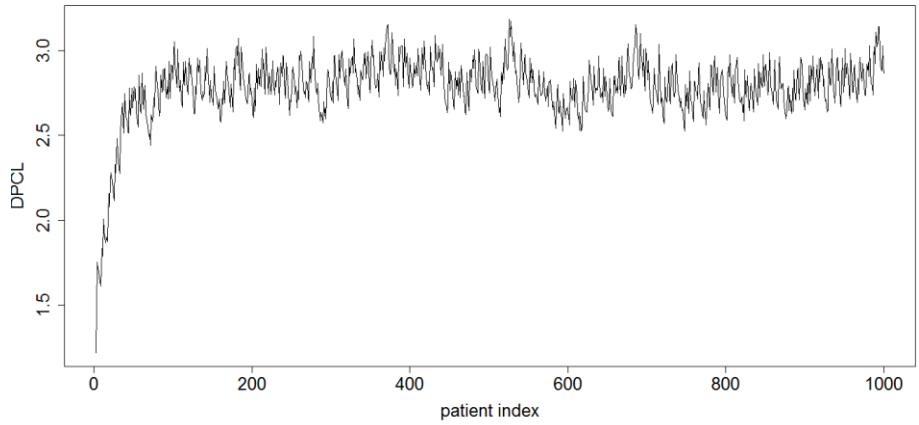


(c) $\alpha_t = \Pr\{C_{t,i} > h_t(0.001)\}$

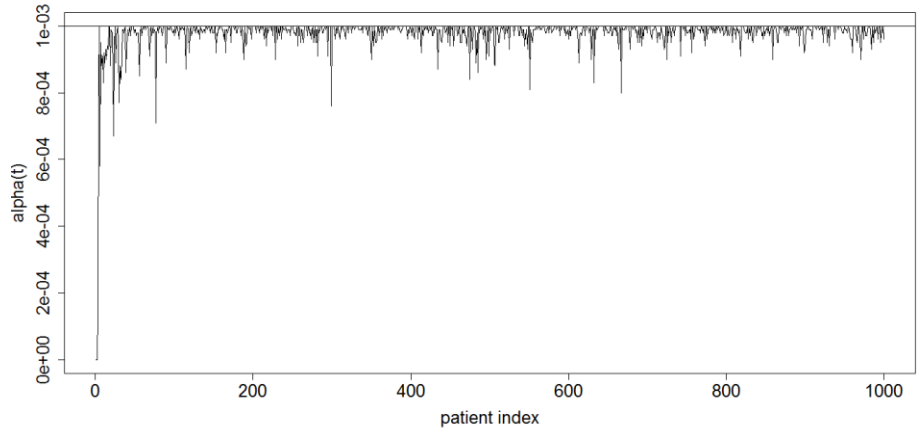
Figure 2.1. (a) Estimated surgical failure rate p_t, (b) DPCLs, and (c) $\alpha_t = \Pr\{C_{t,i} > h_t(0.001)\}$ of the first 50 patients in the first sequence from risk distribution "Surgeon 1" ($\alpha = 0.001$)



(a) Estimated surgical failure rate p_t



(b) DPCLs



(c) $\alpha_t = \Pr\{C_{t,i} > h_t(0.001)\}$

Figure 2.2. (a) Estimated surgical failure rate p_t, (b) DPCLs, and (c) $\alpha_t = \Pr\{C_{t,i} > h_t(0.001)\}$ of the first 1000 patients in the first sequence from risk distribution "Surgeon 1" ($\alpha = 0.001$)

We also compared the DPCLs of two sequences from two risk distributions ("High Risk" and "Low Risk") with the largest difference in mean Parsonnet scores in Figure 2.3. The conditional false alarm rate for both sequences was set to be 0.001. The plot shows that different risk distributions can lead to quite different sets of control limits. Using our approach we were able to obtain similar in-control ARLs (1015.2 and 1006.4) and average conditional false alarm rates of 9.8222E-4 and 9.9257E-4, respectively.

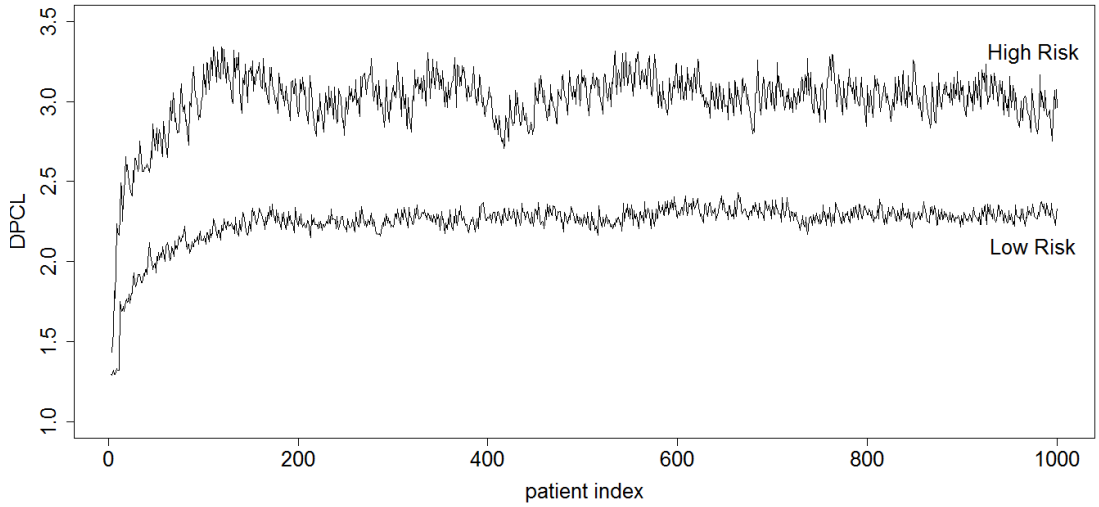


Figure 2.3. DPCLs comparison of two sequences from two different risk distributions ("High Risk" and "Low Risk") (α = 0.001)

If the distributions of risk scores changes over time, as it would for a new surgeon that gains more experience, then our method will result in appropriate limits. This is illustrated in Figure 2.4 where the population shifts from the lower risk patients to the higher risk patients after patient 500. Here α was set as 0.001 and the in-control ARL was estimated to be 1001.60 (S.E. = 3.16), very close to $1/\alpha$.

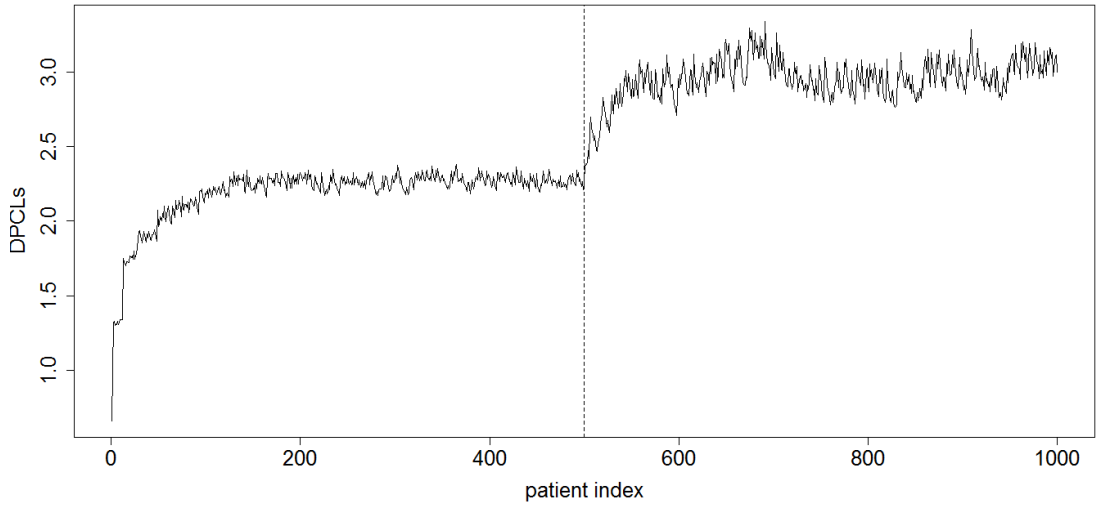


Figure 2.4. DPCLs where the risk distribution shifted from "Low Risk" to "High Risk" after the 500th patient (α = 0.001)

The simulation results show that our method has consistent performance for any patient mix. By using the proposed DPCLs, we are able to obtain a desired in-control run length distribution close to the geometric distribution for any patient sequence from any risk distribution.

4.2. Comparison of constant control limit and DPCLs

From the previous study of Tian *et al*. [21], the upper CUSUM chart in-control ARL for the risk distribution of all scores from 2,218 patients is 7,400.1 applying the constant control limit h^+ = 4.5 under the assumption that the patient population is known. To compare the constant control limit with our proposed DPCLs, we set the false alarm rate at 1/7400.1 = 0.00013513 to obtain the same in-control ARL. The results are shown in Figure 2.5. The dashed line indicates the constant control limit h^+ = 4.5, while the solid line tracks the DPCLs for a particular sequence of scores chosen from the risk distribution. The two types of control limits result in very similar in-

control ARL performance for the risk-adjusted Bernoulli CUSUM chart. The advantage of our method is that we are able to obtain the same in-control ARL and maintain the conditional false alarm rates at the specified level without requiring any information or assumptions about the patient populations. Also, with our approach, it is possible to signal sooner for process changes that occur near the start of the monitoring process. With the DPCLs one could signal at the third patient whereas with the constant control limit a signal could not be given until the eighth patient. Figure 2.5 would lead one to believe that steady-state out-of-control performance would be quite similar.

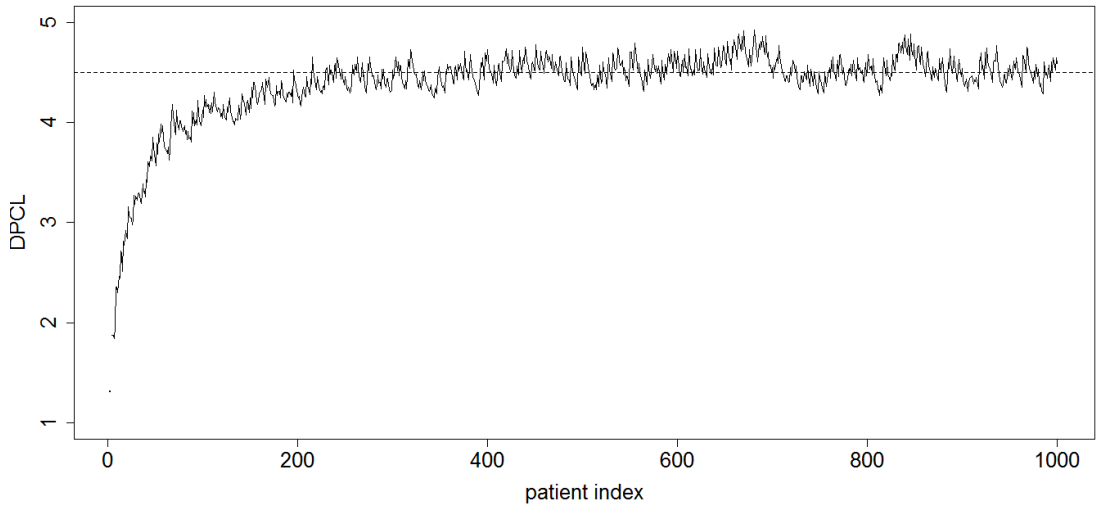


Figure 2.5. Comparison of constant control limit (dashed line) and DPCLs (solid line) for comparable in-control ARLs (α = 0.00013513)

In addition, we compared in Figure 2.6 the conditional false alarm rates for the two charts shown in Figure 2.5. It is quite clear that the conditional false alarm rates of the risk-adjusted CUSUM chart with DPCLs (darker line) were controlled at a specified level α = 1/7400.1 = 0.00013513 while the ones of the chart with the constant control limit (lighter line) varied considerably.

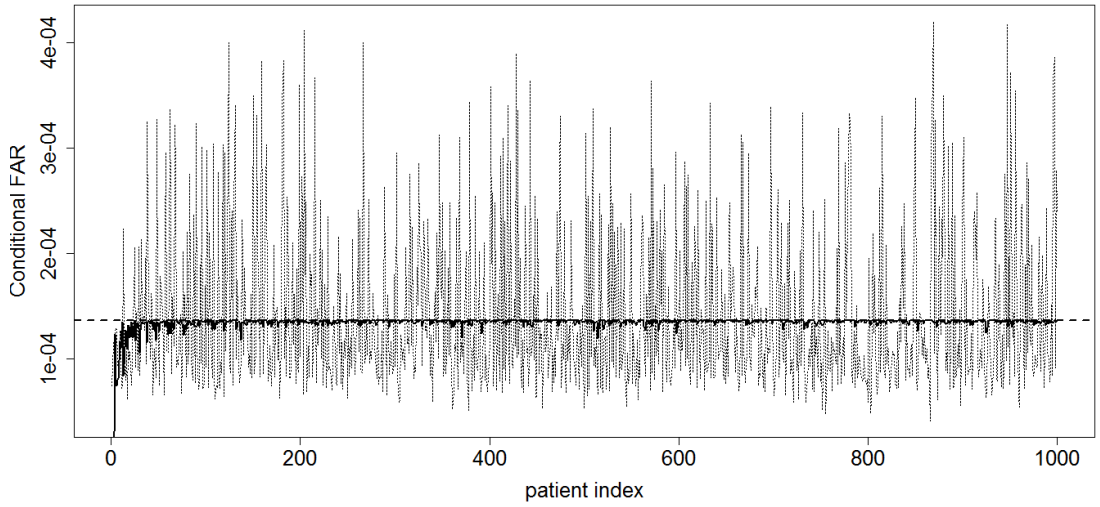


Figure 2.6. Comparison of the conditional false alarm rates (FARs) for the DPCLs control chart (darker line) and the constant limit chart (lighter line) shown in Figure 2.5 (α = 0.00013513)

4.3. Estimated out-of-control performance

We also examined the out-of-control run length performance of the risk-adjusted Bernoulli CUSUM chart with DPCLs compared to the performance of the chart with constant control limits under the assumption that the patient risk distribution is known. Out-of-control performance comparisons are a little complicated, however, by the fact that the two types of charts are based on controlling different aspects of in-control performance.

We used the two charts compared in Section 4.2 for illustration and studied their out-of-control performance. Since the charts were designed to detect a sustained doubling of the odds ratio, the shifted probability of the adverse event was set to be $R_a p_t / (1 - p_t + R_a p_t) = 2p_t / (1 + p_t)$ where p_t is the in-control probability obtained from equation (3.1) for patient *t*. We assumed the odds ratio increased at patient τ = 1, 50, 100, 150,

200, 250, 500, 1000 and 100,000 charts for each method and each change point were simulated to estimate the out-of-control ARLs.

Table 2.5 presents the comparison of out-of-control ARLs of the two types of charts. It is clear that both charts have comparable out-of-control performance under most of the scenarios, except for the earlier change points where the risk-adjusted CUSUM chart with DPCLs could deliver quicker detection. Our results demonstrated the advantage of possibly quicker detection by applying DPCLs instead of constant limits mentioned in Section 4.2. However, it is worth noting that the main purpose of applying DPCLs is to obtain the desired in-control run length distribution and control the conditional false alarm rates rather than to improve the detection ability of the chart [22]. If the risk distribution is unknown or misspecified then the use of a constant control limit will lead to an in-control ARL different from the desired value. In this case, it is not meaningful to compare out-of-control performance.

Table 2.5. Comparison of the estimated out-of-control ARL performance for the two charts compared in Section 4.2 with different τ

Change point τ	Estimated out-of-control ARL (S.E.)	
	RA-CUSUM with DPCLs	RA-CUSUM with $h^+ = 4.5$
1	179.1 (0.44)	206.3 (0.41)
50	181.1 (0.42)	189.6 (0.42)
100	182.3 (0.43)	185.4 (0.42)
150	185.2 (0.43)	187.5 (0.43)
200	190.7 (0.43)	192.7 (0.42)
250	192.9 (0.43)	193.6 (0.42)
500	193.1 (0.43)	192.0 (0.42)
1000	189.2 (0.41)	185.7 (0.40)

5. Discussion

The risk-adjusted Bernoulli CUSUM chart with constant control limits has become a popular tool for monitoring clinical and surgical performance. However, its design requires information or assumptions about patient populations which are often unavailable or inaccurate. Previous studies showed that the effect of varying risk distributions of patients on the in-control performance of risk-adjusted CUSUM charts is significant. Thus, setting fixed control limits for risk-adjusted CUSUM charts in different applications is not suitable in practice. To overcome this disadvantage, we apply dynamic probability control limits to risk-adjusted Bernoulli CUSUM charts. By maintaining the probability of a false alarm at a constant level conditional on no false alarm for previous observations, our charts with DPCLs give desirably consistent in-control performance with approximately geometrically distributed run lengths. The simulation results illustrate that our method does not rely on any information or assumptions about the patient populations or their risk distributions. Thus, the risk-adjusted Bernoulli CUSUM chart with DPCLs is more practical and should be applied to appropriately monitor surgical performance by hospitals and healthcare practitioners. The method is computationally intensive, which will require appropriate software in order to be implemented.

Software to update the DPCLs online for risk-adjusted Bernoulli CUSUM charts is being developed for practical use in applications. Less computationally intensive algorithms to determine DPCLs should be studied as well. For example, a Markov chain model was used to compute the DPCLs for the EWMA charts studied by Shen *et al.* [22]. Finally, our approach could be easily adapted to design the risk-adjusted CUSUM method of Tang *et al.* [27] where there can be more than two possible outcomes for each patient.

References

1. Rogers CA, Reeves BC, Caputo M, Ganesh JS, Bosner RS, Angelini GD. Control chart methods for monitoring cardiac surgical failure and their interpretation. *The Journal of Thoracic and Cardiovascular Surgery* 2004; **128**(6):811-819.

2. Woodall WH, Fogel SL, Steiner SH. The monitoring and improvement of surgical outcome quality. To appear in *Journal of Quality Technology* 2015.

3. Steiner SH, Cook RJ, Farewell VT, Treasure T. Monitoring surgical performance using risk-adjusted cumulative sum charts. *Biostatistics* 2000; **1**(4):441-452.

4. Steiner SH, Cook RJ, Farewell VT. Risk-adjusted monitoring of binary surgical outcomes. *Medical Decision Making* 2001; **21**(3):163-169.

5. Grigg O, Farewell V. An overview of risk-adjusted charts. *Journal of the Royal Statistical Society – Series A* 2004; **167**(3):523-539.

6. Steiner SH. Risk-adjusted monitoring of outcomes in health care. Chapter 14 in *Statistics in Action: A Canadian Outlook*, edited by J. F. Lawless, Chapman and Hall/CRC, 2014; 245-264.

7. Beiles CB, Morton AP. Cumulative sum control charts for assessing performance in arterial surgery. *ANZ Journal of Surgery* 2004; **74**(3):146-151.

8. Harris JR, Forbes, TL, Steiner SH, Lawlor K, Derose G, Harris KA. Risk-adjusted analysis of early mortality after ruptured abdominal aortic aneurysm repair. *Journal of Vascular Surgery* 2005; **42**(3):387-391.

9. Novick RJ, Fox SA, Stitt LW, Forbes TL, Steiner S. Direct comparison of risk-adjusted and non–risk-adjusted CUSUM analyses of coronary artery bypass surgery outcomes. *The Journal of Thoracic and Cardiovascular Surgery* 2006; **132**(2):386-391.

10. Moore R, Nutley M, Cina CS, Motamedi M, Faris P, Abuznadah W. Improved survival after introduction of an emergency endovascular therapy protocol for ruptured abdominal aortic aneurysms. *Journal of Vascular Surgery* 2007; **45**(3):443-450.

11. Bottle A, Aylin P. Intelligent information: A national system for monitoring clinical performance. *Health Services Research* 2008; **43**(1p1):1-31.

12. Morton AP, Clements ACA, Doidge SR, Stackelroth J, Curtis M, Whitby M. Surveillance of healthcare-acquired infections in Queensland, Australia: Data and lessons learned in the first 5 years. *Infection Control and Hospital Epidemiology* 2008; **29**(8):695-701.

13. Collins GS, Jibawi A, McCulloch P. Control charts methods for monitoring surgical performance: A case study from gastro-oesophageal surgery. *European Journal of Surgical Oncology* 2011; **37**(6):473-480.

14. Chen TT, Chung KP, Hu FC, Fan CM, Yang MC. The use of statistical process control (Risk-adjusted CUSUM, risk-adjusted RSPRT and CRAM with prediction limits) for monitoring the outcomes of out-of-hospital cardiac arrest patients rescued by the EMS system. *Journal of Evaluation in Clinical Practice* 2011; **17**(1):71-77.

15. Sherlaw-Johnson C. A method for detecting runs of good and bad clinical outcomes on variable life-adjusted display (VLAD) charts. *Health Care Management Science* 2005; **8**(1):61-65.

16. Sherlaw-Johnson C, Morton A, Robinson MB, Hall A. Real-time monitoring of coronary care mortality: A comparison and combination of two monitoring tools. *International Journal of Cardiology* 2005; **100**(2):301-307.

17. Sherlaw-Johnson C, Wilson P, Gallivan S. The development and use of tools for monitoring the occurrence of surgical wound infections. *Journal of the Operational Research Society* 2007; **58**(2):228-234.

18. Cook DA, Duke G, Hart GK, Pilcher D, Mullany D. Review of the application of risk-adjusted charts to analyze mortality outcomes in critical care. *Critical Care Resuscitation* 2008; **10**(3):239-251.

19. Clinical Practice Improvement Centre. *VLADs for Dummies.* Wiley Publishing Australia Pty Ltd: Milton, Queensland, 2008.

20. Loke CK, Gan FF. Joint monitoring scheme for clinical failures and predisposed risks. *Quality Technology and Quantitative Management* 2012; **9**(1):3-21.

21. Tian WM, Sun HY, Zhang X, Woodall WH. The impact of varying patient populations on the in-control performance of the risk-adjusted CUSUM chart. To appear in *International Journal for Quality in Health Care* 2015.

22. Shen X, Tsung F, Zou C, Jiang W. Monitoring Poisson count data with probability control limits when sample sizes are time-varying. *Naval Research Logistics* 2013; **60**(8):625-636.

23. Margavio TM, Conerly MD, Woodall WH, Drake LG. Alarm rates for quality control charts. *Statistics and Probability Letters* 1995; **24**(3):219-224.

24. Hawkins, DM, Qiu P, Kang CW. The changepoint model for statistical process control. *Journal of Quality Technology* 2003; **35**(4):355-366.

25. Gombay E, Hussein AA, Steiner SH. Monitoring binary outcomes using risk-adjusted charts: a comparative study. *Statistics in Medicine* 2011; **30**(23):2815-2816.

26. Parsonnet V, Dean D, Berstein AD. A method of uniform stratification of risk for evaluating the results of surgery in acquired adult heart disease. *Circulation* 1989; **79**(6):3-12.

27. Tang X, Gan FF, Zhang L. Risk-adjusted cumulative sum charting procedure based on multi-responses. To appear in *Journal of the American Statistical Association* 2015.

Chapter 3

Dynamic probability control limits for lower and two-sided risk-adjusted Bernoulli CUSUM charts

ABSTRACT

Due to its advantages of design, performance and effectiveness in reducing the effect of patients' prior risks, the risk-adjusted Bernoulli cumulative sum (CUSUM) chart developed by Steiner *et al.* (2000) is widely applied to monitor clinical and surgical outcome performance. In practice, it is beneficial to obtain evidence of improved surgical performance using the lower risk-adjusted Bernoulli CUSUM charts. However, it had been shown that the in-control performance of the charts with constant control limits varies considerably for different patient populations. In our study, we apply the dynamic probability control limits (DPCLs) developed for the upper risk-adjusted Bernoulli CUSUM charts by Zhang and Woodall (2015) to the lower and two-sided charts and examine their in-control performance. The simulation results demonstrate that the in-control performance of the lower risk-adjusted Bernoulli CUSUM charts with DPCLs can be controlled for different patient populations because these limits are determined for each specific sequence of patients. In addition, practitioners could also run upper and lower risk-adjusted Bernoulli CUSUM charts with DPCLs side by side simultaneously and obtain desired in-control performance for the two-sided chart for any particular sequence of patients for a surgeon or hospital.

Keywords: average run length (ARL); false alarm rate; statistical process monitoring; surgical outcome quality.

1. Introduction

It is of increasing interest to monitor the quality of healthcare performance, with particular emphasis on the quality of surgical outcome performance. It is important to detect any deterioration of surgical performance in a timely manner in order to avoid adverse consequences. Moreover, to obtain evidence of improved surgical performance is certainly beneficial for healthcare practitioners and organizations as well.

Due to the considerable variability of pre-operative risks in different patient populations, the standard Bernoulli cumulative sum (CUSUM) chart of Reynolds and Stoumbos (1999) is not appropriate for monitoring surgical outcome quality since it doesn't consider the heterogeneity among patients. To alleviate this problem, Steiner *et al.* (2000) and Steiner *et al.* (2001) developed the risk-adjusted Bernoulli CUSUM chart that adjusts for each patient's pre-operative risk of surgical failure through the use of a logistic regression model and a likelihood-ratio-based scoring method to obtain the monitoring statistics. The risk-adjusted Bernoulli CUSUM chart has been shown to be suited for settings where there is a mix of patients with various pre-operative risks and therefore has been recommended by many and applied by a number of practitioners for various applications. For example, Grigg and Farewell (2004), Woodall (2006), Cook *et al.* (2008), Steiner (2014) and Woodall *et al.* (2015) provided the overviews of risk-adjusted monitoring. Beiles *et al.* (2004), Harris *et al.* (2005), Sherlaw-Johnson (2005), Sherlaw-Johnson *et al.* (2005), Novick *et al.* (2006), Moore *et al.* (2007), Sherlaw-Johnson *et al.* (2007), Bottle and Aylin (2008), Morton *et al.* (2008), Collins *et al.* (2011) and Chen *et al.* (2011) used the risk-adjusted CUSUM charts to assess or monitor clinical outcome performance for various applications. However, several researchers have brought up issues about the effect of different

risk distributions on the performance of risk-adjusted Bernoulli CUSUM charts (See Rogers *et al.* (2004), Steiner *et al.* (2001), Loke and Gan (2012) and Tian *et al.* (2015)). The in-control average run lengths (ARLs) of risk-adjusted CUSUM charts with the same risk adjustment model and constant control limits can vary by a factor of ten for the highest and lowest risk patient populations. Therefore, Steiner *et al.* (2001) recommended that the control limit of the monitoring procedure be adjusted if the patients mix changes substantially.

Zhang and Woodall (2015) recently developed a simulation-based procedure to determine the dynamic probability control limits (DPCLs) for risk-adjusted Bernoulli CUSUM charts based on the method of Shen *et al.* (2013), who considered an application involving the exponentially weighted moving average (EWMA) chart to monitor Poisson count data with time-varying population sizes. The concept in non-dynamic applications was previously used by Margavio *et al.* (1995) and Hawkins *et al.* (2003). By maintaining the conditional false alarm rate at a constant value given that there are no false alarms for previous observations, one can design the risk-adjusted CUSUM chart with DPCLs which result in approximately geometrically distributed in-control run lengths with a desired in-control ARL for any sequence of patients. The simulation results of applying DPCLs to upper risk-adjusted Bernoulli CUSUM charts showed that the method does not require any assumptions about the patients' risk distribution. Therefore, using DPCLs instead of constant control limits overcomes the major disadvantage of the risk-adjusted Bernoulli CUSUM chart and is more practical to use for monitoring surgical outcome quality. Moreover, patient populations can change over time and are not easy to estimate accurately.

In our study, we apply DPCLs to the lower risk-adjusted Bernoulli CUSUM charts and examine the in-control performance of the lower and the two-sided charts. The remainder of this

paper is organized as follows. The procedure to determine the DPCLs for risk-adjusted Bernoulli CUSUM charts and our simulation settings are explained in Section 2. This is followed by the simulation results of the lower charts with DPCLs and some examples for illustration in Section 3. Next, the in-control performance of two-sided risk-adjusted Bernoulli CUSUM charts with DPCLs is examined in Section 4. Finally, we provide some discussion and conclusions in Section 5.

2. Methods and simulation settings

2.1. Methods

The upper tabular CUSUM statistics (C_0^+) and lower tabular CUSUM statistics (C_0^-) can be written as follows,

$$\begin{cases} C_t^+ = \max\left(0, C_t^+ + W_t^+\right) \\ C_t^- = \min\left(0, C_t^- - W_t^-\right) \end{cases}, t = 1, 2, 3, \ldots \quad (1)$$

where $C_0^+ = C_0^- = 0$ and W_t^+ and W_t^- represent the weight assigned to the result for the t^{th} individual for the upper and lower chart, respectively. The risk-adjusted CUSUM chart developed by Steiner *et al.* (2000) is designed to monitor for an odds ratio change from R_0 to R_a, where R_0 is usually set to 1 to reflect current expected surgical performance, and set to a value $R_a > R_0$ for detecting performance deterioration in the upper chart, and to a value $R_a < R_0$ to detect process improvement in the lower chart. For the risk-adjusted Bernoulli CUSUM charts, the weights are calculated using:

$$W_t = \begin{cases} \log\left(\dfrac{1 - p_t + R_0 p_t}{1 - p_t + R_a p_t}\right) & \text{if } y_t = 0, \\ \log\left[\dfrac{\left(1 - p_t + R_0 p_t\right) R_a}{\left(1 - p_t + R_a p_t\right) R_0}\right] & \text{if } y_t = 1. \end{cases} \quad (2)$$

where p_t is the surgical failure rate for each patient determined by assessing the patient's pre-operative risk and applying a method such as a logistic regression model based on Parsonnet scores (Parsonnet *et al.* (1989)). We have y_t = 1 if patient *t* experiences the adverse event of

interest, such as death or a surgical site infection within a specified time period following surgery, and y_t = 0 otherwise. The chart signals when $C_t^+ > h^+$ or $C_t^- < h^-$ which indicates that there has been either deterioration or improvement in the surgical performance, respectively. The control limits h^+ and h^- are set to yield a suitably large in-control ARL value when there are no changes in the odds ratio of failure R_0.

Zhang and Woodall (2015) applied dynamic probability control limits (DPCLs) to the upper risk-adjusted Bernoulli CUSUM charts. The method of determining DPCLs for the lower risk-adjusted Bernoulli CUSUM charts is similar to the one for the upper CUSUM charts used to detect process deterioration. The main idea is to maintain the conditional probability of obtaining a false alarm at a constant level from patient to patient given there are no false alarms for the previous patients. Particularly, for a lower risk-adjusted Bernoulli CUSUM chart, the DPCLs $\boldsymbol{h}^-(\alpha) = \left(h_1^-(\alpha),\ h_2^-(\alpha),\ ...,\ h_k^-(\alpha),\ ...\right)$ are designed to satisfy to the extent possible the following equations:

$$\begin{cases} \Pr\left(C_1^- < h_1^-(\alpha) \middle| S_1\right) = \alpha, \\ \Pr\left(C_t^- < h_t^-(\alpha) \middle| C_k^- \geq h_k^-(\alpha),\ k = 1,\ ...,\ t-1,\ S_t\right) = \alpha, \text{ for } t = 2,\ 3,\ \ldots, \end{cases} \tag{3}$$

where S_1, S_2, … is the sequence of observed risk scores, such as Parsonnet scores, and α is the predetermined conditional false alarm rate. For each patient, we calculate the observed risk-adjusted CUSUM statistic $C_t^- = \min\left(0, C_{t-1}^- - W_t^-\right)$, where W_t^- is computed using Equation (2) based on the observed outcome y_t and the failure rate p_t. The chart signals when $C_t^- < h_t^-(\alpha)$. From Equation (3), we can clearly see that the in-control run length is approximately

geometrically distributed with parameter α. Thus, it is expected to find the in-control ARL with use of the DPCLs, $\boldsymbol{h}^{-}(\alpha) = \left(h_1^{-}(\alpha),\ h_2^{-}(\alpha),\ ...,\ h_k^{-}(\alpha),\ ...\right)$, to be close to $1/\alpha$.

The computational procedure to obtain the DPCLs for lower risk-adjusted Bernoulli CUSUM chart is outlined below. At the beginning, consider the first patient with some risk score S_1. Under expected performance, the estimated rate p_1 can be calculated from the risk-adjustment model. Next, we generate N random variables $Y_{1,i}$ which are independent and identically distributed (i.i.d.) Bernoulli random variables with probability p_1, i = 1, 2, …, N, where N is a sufficiently large number. Then we calculate the lower risk-adjusted CUSUM statistics $C_{1,1}^{-}$, $C_{1,2}^{-}$, $C_{1,3}^{-}$, …, $C_{1,N}^{-}$ corresponding to the $Y_{1,i}$'s, using equations (1) and (2), sort the N CUSUM statistics in ascending order $C_{1,(1)}^{-}$, $C_{1,(2)}^{-}$, $C_{1,(3)}^{-}$, …, $C_{1,(N)}^{-}$ and take the $N' = \lfloor N \cdot \alpha \rfloor + 1$ smallest CUSUM statistic $C_{1,(N')}^{-}$ as the approximated DPCL $h_1^{-}(\alpha)$. Note that the CUSUM statistic can only take two values at t = 1 (i.e. the first patient) due to the binary nature of Bernoulli variables. Note also that for some patients there may not be a control limit such that the approximated conditional false alarm rate is α or below. In these cases, there is no control limit used and no possibility of the chart signaling.

Then, for patient 2 with risk score S_2, we generate N i.i.d. Bernoulli random values $Y_{2,i}$ with probability p_2, where p_2 is calculated from our logistic regression model. Again, N corresponding lower risk-adjusted CUSUM statistics $C_{2,i}^{-}$, i = 1, 2, …, N are computed from Equations (1) and (2) with randomly chosen values from a vector of $C_{1,i}^{-}$ values which are greater than or equal to $h_1^{-}(\alpha)$. Then we sort the N $C_{2,i}^{-}$ values in ascending order and take the

$N' = \lfloor N \cdot \alpha \rfloor + 1$ smallest CUSUM statistic $C^-_{2,(N')}$ as the DPCL $h^-_2(\alpha)$. For the next patient, we keep the CUSUM statistics which are greater than or equal to $h^-_2(\alpha)$ and repeat the simulation procedure to obtain $h^-_3(\alpha)$. The procedure is continued iteratively to obtain the following control limits. The algorithmic form of the simulation procedure to obtain the control limit for patient *t* (*t* = 1, 2, 3, …) can be summarized as follows:

1) Generate *N* Bernoulli random variables $Y_{t,i}$ (*i* = 1, 2, …, *N*) with in-control failure rate p_t obtained from risk-adjustment model and *N* CUSUM statistics $C^-_{t,i} = \min\left(0, C^-_{t-1,j} - W^-_{t,i}\right)$ (*j* = 1, 2, …, *N*) using Equations (1) and (2) accordingly, where $C^-_{t-1,j}$ is randomly selected from a vector of CUSUM values $\boldsymbol{C}^{-\prime}_{t-1} = \left\{C^-_{t-1,i} \text{ such that } C^-_{t-1,i} \geq h^-_{t-1}(\alpha)\right\}$.

2) Sort the *N* CUSUM statistics in ascending order $C^-_{t,(1)}$, $C^-_{t,(2)}$, $C^-_{t,(3)}$, …, $C^-_{t,(N)}$ and take the $N' = \lfloor N \cdot \alpha \rfloor + 1$ smallest CUSUM statistic $C^-_{t,(N')}$ as the approximated DPCL $h^-_t(\alpha)$ if $\Pr\left\{C^-_{t,i} < C^-_{t,(N')}\right\} \neq 0$.

3) Calculate $C^-_t = \min\left(0, C^-_{t-1} - W^-_t\right)$ based on the surgical outcome and risk score of patient *t*. If $C^-_t < h^-_t(\alpha)$, an out-of-control signal is issued. Otherwise, go back to step 1.

Due to the discreteness of Bernoulli random variables, for each patient *t* = 1, 2, … , $\Pr\left\{C^-_{t,i} < h^-_t(\alpha)\right\}$ (denoted by α_t) is always controlled to be less than or equal to α. As the control limit determination procedure progresses, an increasing number of different CUSUM statistic values will be generated for each patient. After a number of patients, we will observe

more $\alpha_t = \Pr\left\{C_{t,i}^- < h_t^-(\alpha)\right\}$ values close to α. However, as shown in Section 3, the observed conditional false alarm rate does not converge to α as the number of patients increases.

Figure 3.1 shows an example of DPCLs obtained for the first 1000 patients in a sequence randomly chosen from the patient population "Surgeon 1" by applying the proposed procedure. Here the conditional false alarm rate α was set as 0.0005. We can see that the probability control limits vary dynamically from patient to patient. Note also that for some of the patients there are no control limits due to the fact that $\Pr\left\{C_{t,i}^- < C_{t,(N')}^-\right\} = 0$.

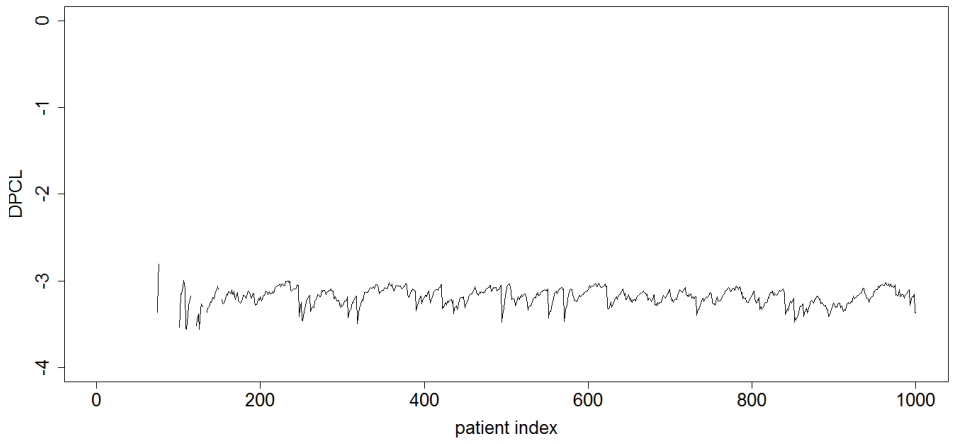


Figure 3.1. DPCLs of the first 1000 patients in a sequence from risk distribution "Surgeon 1" (α = 0.0005) for lower risk-adjusted Bernoulli CUSUM chart

2.2. Simulation settings

Tian *et al.* (2015) found that the in-control ARL of the Bernoulli risk-adjusted CUSUM chart with constant control limits can vary by a factor of two for different realistic patient populations. The in-control ARL increases as the mean risk score of the patient population decreases.

The data set of patients used in the studies of Steiner *et al.* (2000) and Zhang and Woodall (2015) was based on 6994 surgeries from a UK surgical center over a 7-year period. In our simulations for the lower risk-adjusted Bernoulli CUSUM charts, we used the same Parsonnet score sequences examined by Zhang and Woodall (2015) for the upper CUSUM charts and the same logistic regression model,

$$\text{logit}(p_t) = -3.68 + 0.077 X_t,$$

where X_t is the Parsonnet score of patient t and p_t is the probability of death within thirty days following surgery. The model was fitted using the first two years of data containing 2218 patients' records which were treated as in-control group. The Parsonnet score sequences were randomly selected with replacement from each of the following five different Phase I patient populations represented by different Parsonnet score distributions defined by Tian *et al.* (2015):

6) All: 2,218 scores for all patients (Mean = 8.9026),
7) High Risk: the highest 50% of the scores (Mean = 16.4813),
8) Low Risk: the lowest 50% of the scores (Mean = 2.0541),
9) Surgeon 1: 565 scores for all of surgeon 1's patients (Mean = 11.2513),
10) Surgeon 6: 474 scores for all of surgeon 6's patients (Mean = 5.5591).

For each sequence of scores, we calculated the DPCLs applying the algorithm outlined in Section 2.1. Several conditional false alarm rates α were examined. Also, N, the number of CUSUM statistics simulated, was set to be a suitably large number. The risk-adjusted CUSUM statistics were obtained using Equations (1) and (2) with $R_0 = 1$ and $R_a = 0.5$. Moreover, the observed conditional false alarm rate $\alpha_t = \Pr\left\{C_{t,i}^- < h_t^-(\alpha)\right\}$ for each patient t and the proportion

of no control limits, P_{No}, for each sequence were recorded as a check for accuracy of the procedure. After we obtained the DPCLs for the specific sequence of scores, 100,000 control charts were simulated to estimate the in-control ARL (ARL_0). Also, the 10^{th}, 25^{th}, 50^{th}, 75^{th}, 90^{th} percentiles of the run length ($Q_{0.10}$, $Q_{0.25}$, $Q_{0.50}$, $Q_{0.75}$, and $Q_{0.90}$) and standard deviation of the run length (SDRL) were estimated in order to compare with those of the geometric distribution.

3. Results

Similar to the study on upper CUSUM charts with DPCLs, we first set the conditional false alarm rate α to be 0.005 and N = 100,000. The simulation results are listed in Table 3.1 for five sequences from each risk distribution along with the parameters of the geometric distribution with p = 0.005 for comparison.

Table 3.1. Estimated in-control performance of lower risk-adjusted Bernoulli CUSUM charts with DPCLs (α = 0.005)

Risk distribution	Sequence index	$\widehat{\text{ARL}_0}$(S.E.)	$\widehat{\text{SDRL}}$	$Q_{0.10}$	$Q_{0.25}$	$Q_{0.50}$	$Q_{0.75}$	$Q_{0.90}$	$\overline{\alpha_t}$	P_{No}
All	1	381.8(0.98)	309.2	113	172	295	502	787	4.3944E-3	0.2790
	2	396.1(1.07)	338.1	92	150	290	525	834	4.3880E-3	0.2787
	3	378.3(1.00)	314.7	108	150	285	494	802	4.3893E-3	0.2801
	4	389.0(1.00)	317.7	105	169	311	510	800	4.3920E-3	0.2833
	5	381.9(1.00)	317.0	103	160	281	503	798	4.4027E-3	0.2812
High Risk	1	305.8(0.80)	253.5	78	124	221	408	640	4.4203E-3	0.1146
	2	304.6(0.82)	258.3	80	122	225	397	642	4.4317E-3	0.1195
	3	312.9(0.81)	256.8	85	131	233	406	647	4.4333E-3	0.1174
	4	297.2(0.79)	248.7	76	118	217	396	623	4.4317E-3	0.1205
	5	312.5(0.80)	252.7	92	133	230	415	634	4.4194E-3	0.1162
Low Risk	1	429.0(0.81)	256.1	206	247	345	527	761	4.0836E-3	0.0827
	2	446.5(0.83)	261.6	220	262	365	544	792	4.0917E-3	0.0889
	3	414.8(0.84)	265.5	192	229	330	507	769	4.0758E-3	0.0869
	4	426.0(0.82)	259.3	201	245	347	518	756	4.0707E-3	0.0857
	5	429.4(0.81)	257.1	204	249	340	530	759	4.0782E-3	0.0845
Surgeon 1	1	355.0(0.93)	292.9	92	140	265	467	731	4.3899E-3	0.2134
	2	367.8(0.89)	280.3	117	163	280	473	722	4.3892E-3	0.2042
	3	363.5(0.92)	291.3	100	142	270	496	756	4.3978E-3	0.2143
	4	333.4(0.90)	284.3	83	140	250	445	701	4.3914E-3	0.2130
	5	337.0(0.92)	289.7	90	133	242	434	705	4.3853E-3	0.2097
Surgeon 6	1	431.3(1.07)	338.6	136	209	316	550	899	4.3931E-3	0.3447
	2	443.5(1.08)	342.3	153	209	348	570	907	4.3790E-3	0.3328
	3	437.1(0.98)	311.1	169	218	342	544	820	4.3762E-3	0.3287
	4	450.1(1.05)	333.5	144	221	347	575	880	4.3769E-3	0.3335
	5	451.8(1.03)	326.6	176	218	348	573	857	4.3947E-3	0.3413
Geometric	/	200 (/)	200	21	57	138	276	459	/	

Secondly, we considered a lower conditional false alarm rate of $\alpha = 0.001$ with N = 100,000. The results are summarized in Table 3.2 for five sequences from each risk distribution and compared with the corresponding parameters of the geometric distribution with $p = 0.001$.

Table 3.2. Estimated in-control performance of lower risk-adjusted Bernoulli CUSUM charts with DPCLs ($\alpha = 0.001$)

Risk distribution	Sequence index	$\widehat{\text{ARL}_0}$(S.E.)	$\widehat{\text{SDRL}}$	$Q_{0.10}$	$Q_{0.25}$	$Q_{0.50}$	$Q_{0.75}$	$Q_{0.90}$	$\overline{\alpha_t}$	P_{No}
All	1	1299.9(3.73)	1179.9	250	448	943	1778	2854	8.8123E-4	0.0621
	2	1343.2(3.85)	1218.6	276	487	959	1810	2919	8.8326E-4	0.0681
	3	1330.8(3.71)	1174.3	261	497	974	1800	2861	8.8308E-4	0.0652
	4	1319.0(3.79)	1198.2	255	462	959	1770	2878	8.8236E-4	0.0647
	5	1324.0(3.77)	1191.1	273	478	938	1805	2878	8.8171E-4	0.0714
High Risk	1	1124.3(3.41)	1079.8	169	371	790	1525	2499	9.1651E-4	0.0101
	2	1182.2(3.43)	1085.6	209	407	852	1604	2608	9.1857E-4	0.0081
	3	1161.4(3.41)	1079.2	204	395	828	1570	2568	9.1769E-4	0.0082
	4	1177.6(3.45)	1090.0	201	410	840	1595	2602	9.1726E-4	0.0093
	5	1191.3(3.46)	1095.0	212	415	841	1615	2628	9.1716E-4	0.0090
Low Risk	1	1719.4(4.53)	1433.6	447	685	1294	2297	3594	8.2181E-4	0.1959
	2	1675.0(4.57)	1446.4	419	654	1199	2243	3578	8.2299E-4	0.1924
	3	1673.5(4.54)	1435.2	396	619	1250	2248	3471	8.2460E-4	0.2004
	4	1740.4(4.75)	1501.9	361	706	1268	2324	3650	8.2617E-4	0.2085
	5	1667.8(4.61)	1459.0	387	609	1212	2231	3533	8.2341E-4	0.1896
Surgeon 1	1	1223.0(3.54)	1119.2	232	427	880	1649	2657	8.9672E-4	0.0342
	2	1229.9(3.62)	1143.3	210	425	876	1660	2732	8.9840E-4	0.0319
	3	1260.3(3.60)	1139.7	240	453	912	1713	2770	8.9630E-4	0.0296
	4	1212.7(3.59)	1135.7	202	399	871	1637	2686	8.9574E-4	0.0312
	5	1253.9(3.62)	1144.5	233	433	894	1722	2748	8.9621E-4	0.0384
Surgeon 6	1	1536.6(4.26)	1348.0	316	590	1118	2058	3308	8.5667E-4	0.1408
	2	1529.6(4.29)	1356.9	326	557	1108	2041	3305	8.5407E-4	0.1385
	3	1503.7(4.19)	1324.9	327	566	1099	2019	3213	8.5558E-4	0.1409
	4	1467.2(4.10)	1296.3	337	542	1069	1931	3124	8.5691E-4	0.1530
	5	1624.5(4.28)	1352.4	377	656	1237	2164	3390	8.5606E-4	0.1408
Geometric	/	1000 (/)	1000	105	287	692	1385	2301	/	

Moreover, we decreased these conditional false alarm rates by a factor of 10 and examined the in-control performance of lower risk-adjusted Bernoulli CUSUM charts with DPCLs. Here we set N = 1,000,000 and implemented the same simulation procedures explained in Section 2.1 for each sequence. The estimated in-control performance is summarized in Table 3.3 and Table 3.4, for α = 0.0005 and α = 0.0001, respectively.

Table 3.3. Estimated in-control performance of lower risk-adjusted Bernoulli CUSUM charts with DPCLs (α = 0.0005)

Risk distribution	Sequence index	$\widehat{\text{ARL}}_0$(S.E.)	$\widehat{\text{SDRL}}$	$Q_{0.10}$	$Q_{0.25}$	$Q_{0.50}$	$Q_{0.75}$	$Q_{0.90}$	$\overline{\alpha_t}$	P_{No}
All	1	2258.0(6.65)	2101.5	384	766	1623	3066	4976	4.7627E-4	0.00332
	2	2268.8(6.64)	2099.2	378	776	1633	3089	4981	4.7612E-4	0.00384
High Risk	1	2160.2(6.48)	2050.7	315	694	1533	2959	4848	4.8411E-4	0.00044
	2	2168.0(6.50)	2056.6	327	703	1534	2961	4852	4.8415E-4	0.00055
Low Risk	1	2413.1(6.74)	2130.3	513	892	1753	3246	5176	4.6938E-4	0.00141
	2	2415.6(6.72)	2126.2	510	897	1761	3264	5191	4.6938E-4	0.00142
Surgeon 1	1	2202.3(6.60)	2088.2	337	721	1556	3004	4924	4.7930E-4	0.00079
	2	2218.7(6.63)	2095.7	339	724	1585	3029	4952	4.7937E-4	0.00075
Surgeon 6	1	2378.0(6.92)	2187.8	429	819	1699	3228	5250	4.7197E-4	0.03319
	2	2390.8(6.94)	2193.1	423	846	1725	3230	5263	4.7189E-4	0.03190
Geometric	/	2000 (/)	2000	210	575	1385	2771	4604	/	/

Table 3.4. Estimated in-control performance of lower risk-adjusted Bernoulli CUSUM charts with DPCLs (α = 0.0001)

Risk distribution	Sequence index	$\widehat{\text{ARL}}_0$(S.E.)	$\widehat{\text{SDRL}}$	$Q_{0.10}$	$Q_{0.25}$	$Q_{0.50}$	$Q_{0.75}$	$Q_{0.90}$	$\overline{\alpha_t}$	P_{No}
All	1	11128.9(34.46)	10896.0	1379	3400	7763	15366	25344	9.0984E-5	0.00421
	2	11194.0(34.74)	10986.8	1400	3397	7839	15426	25478	9.0991E-5	0.00361
High Risk	1	10504.0(33.05)	10450.2	1156	3042	7284	14592	24247	9.4331E-5	0.00058
	2	10697.9(33.39)	10557.9	1243	3135	7493	14778	24332	9.4343E-5	0.00063
Low Risk	1	13611.8(42.23)	13353.7	1781	4169	9515	18626	30865	8.3476E-5	0.11614
	2	13492.2(41.82)	13224.4	1684	4127	9457	18451	30593	8.3580E-5	0.11491
Surgeon 1	1	10897.7(33.86)	10706.2	1268	3233	7671	15061	24852	9.2613E-5	0.00142
	2	10924.1(34.27)	10836.8	1282	3311	7572	15025	24908	9.2612E-5	0.00155
Surgeon 6	1	11872.0(36.74)	11618.2	1504	3564	8266	16380	27116	8.7713E-5	0.02074
	2	11926.7(36.72)	11611.4	1526	3688	8267	16359	27101	8.7729E-5	0.02344
Geometric	/	10000 (/)	10000	1053	2876	6931	13862	23024	/	/

From Tables 3.1 – 3.4, we can see that the in-control ARL performance for the lower risk-adjusted Bernoulli CUSUM charts with DPCLs is affected by the average observed conditional false alarm rate $\overline{\alpha_t}$ and the proportion of no lower control limits P_{No}. The run length distribution gets closer to the theoretical geometric distribution as $\overline{\alpha_t}$ gets closer to α and P_{No} gets closer to 0. We have a greater proportion of time with no lower limit for the larger values of α. For the upper risk-adjusted CUSUM chart, there will be upper control limits for all but the first few patients. Since most patients have small probabilities of mortality within 30 days after surgery, it is more difficult to detect process improvement than process deterioration while maintaining the desired conditional false alarm rate α. Having lower risk patients leads to a greater number of patients for which there are no lower control limits.

We used sequences chosen from the five risk distributions to examine the observed conditional false alarm rates $\alpha_t = \Pr\left\{C_{t,i}^- < h_t^-(\alpha)\right\}$ more closely. Here N is set to be 1,000,000 for all scenarios. The observed conditional false alarm rates of the first 50,000 patients in the sequences from the five risk distributions with α = 0.005, 0.001, 0.0005, 0.0001 are plotted in Figures 3.2 – 3.6, respectively. If $\alpha_t = 0$, then there is no lower control limit at patient t. We notice that, in general, the proportion of no control limits decreases as the conditional false alarm rate decreases except for the "Low Risk" distribution. The outlying pattern may be due to the fact that N is not large enough for that scenario.

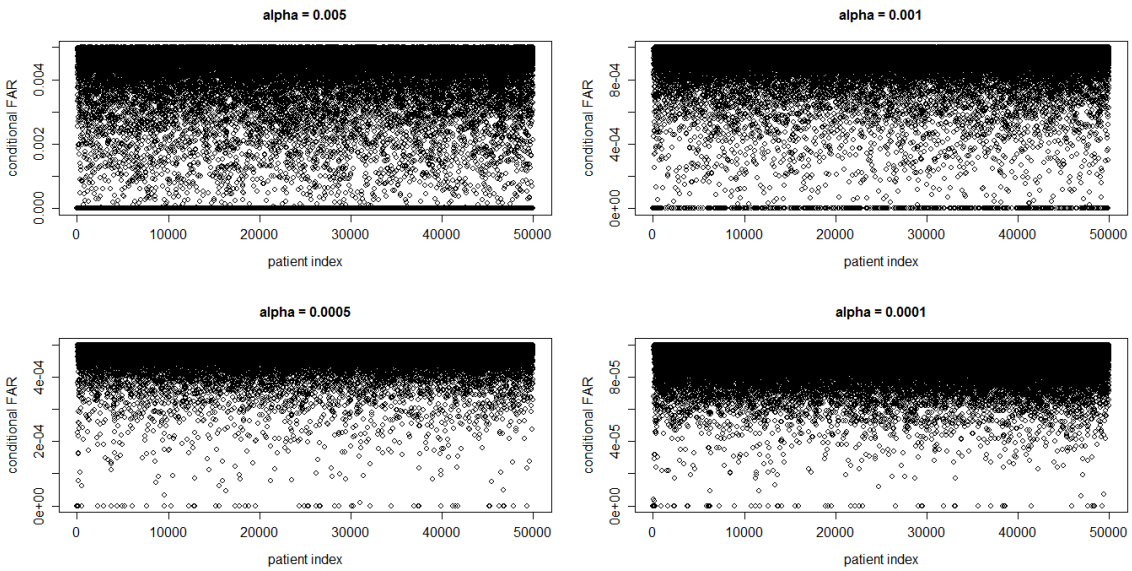


Figure 3.2. Conditional false alarm rates of the first 50,000 patients in a chosen sequence from risk distribution "All" with α = 0.005, 0.001, 0.0005, 0.0001

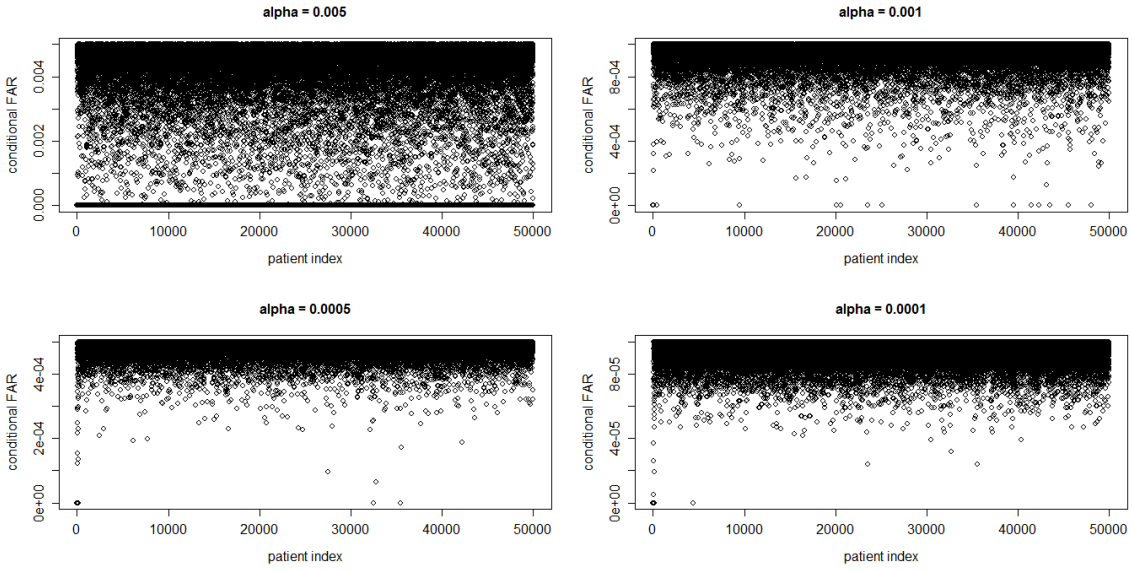


Figure 3.3. Conditional false alarm rates of the first 50,000 patients in a chosen sequence from risk distribution "High Risk" with α = 0.005, 0.001, 0.0005, 0.0001

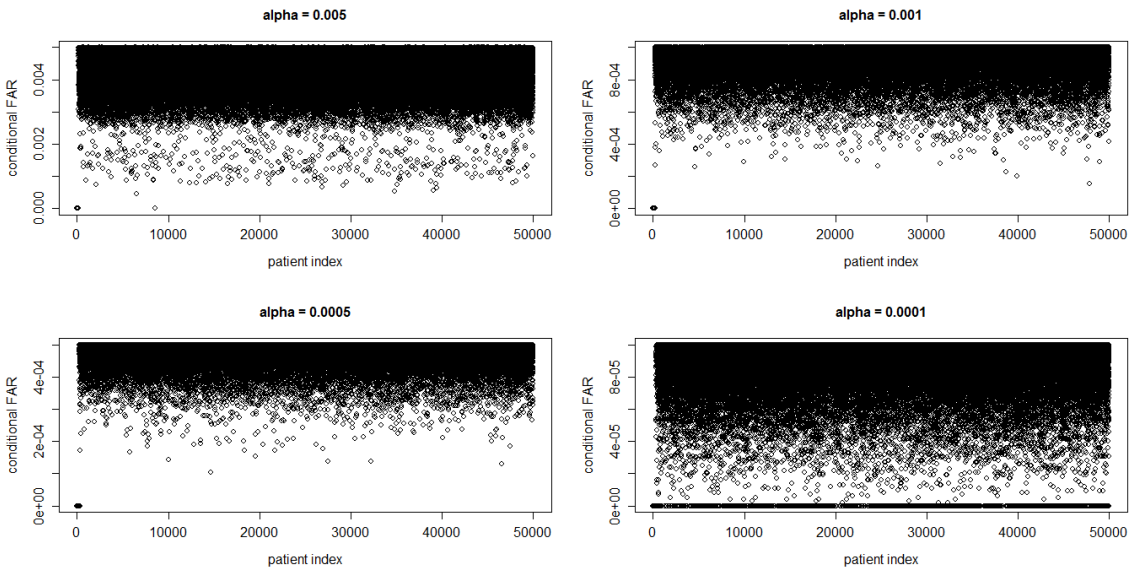


Figure 3.4. Conditional false alarm rates of the first 50,000 patients in a chosen sequence from risk distribution "Low Risk" with α = 0.005, 0.001, 0.0005, 0.0001

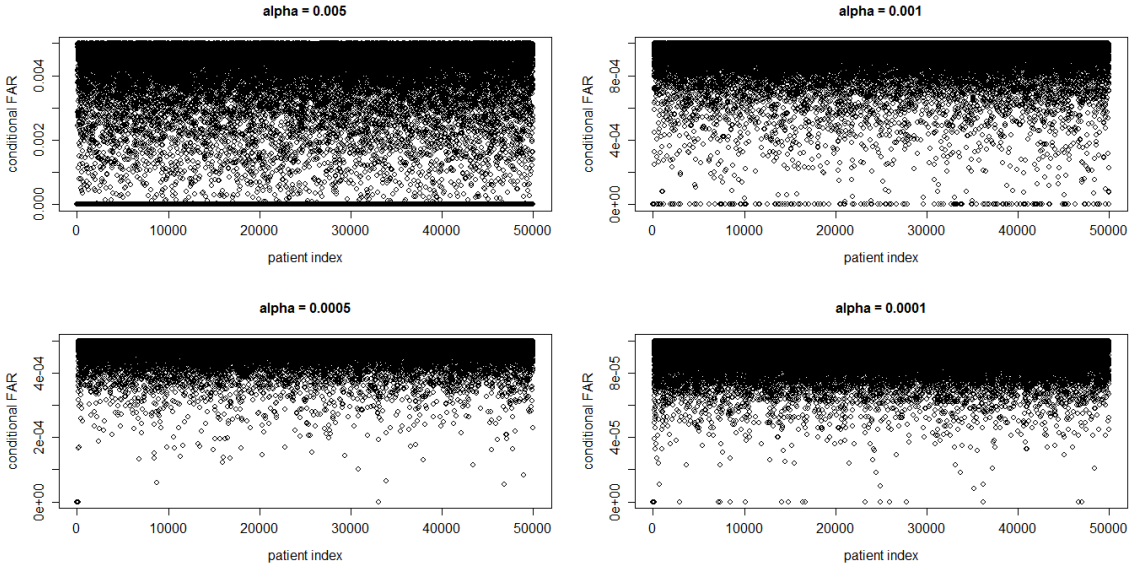


Figure 3.5. Conditional false alarm rates of the first 50,000 patients in a chosen sequence from risk distribution "Surgeon 1" with α = 0.005, 0.001, 0.0005, 0.0001

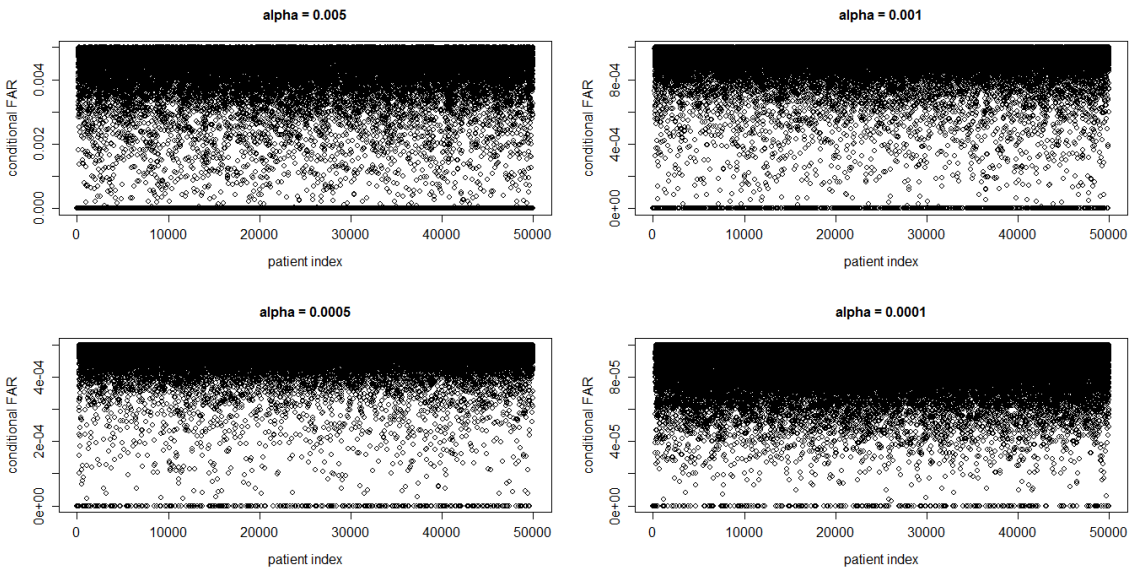


Figure 3.6. Conditional false alarm rates of the first 50,000 patients in a chosen sequence from risk distribution "Surgeon 6" with α = 0.005, 0.001, 0.0005, 0.0001

We further increased N to 1,000,000 for α = 0.005, 0.001 and to 5,000,000 for α = 0.0005, 0.0001 to check whether the value of N would have some effect on the performance of the DPCLs for the same sequences of patients. The results are summarized in Tables 3.5 – 3.8. Compared with the results in Tables 3.1 – 3.4, it is clear that larger value of N can lower the proportion of no control limits P_{No} and increase the average observed conditional false alarm rates $\overline{\alpha_t}$, therefore making the in-control run length distribution closer to the geometric distribution with parameter α.

Despite the deviation from the theoretical geometric distribution, the in-control performance of the charts with DPCLs for different risk distributions varies much less than the chart with constant control limits. Thus, DPCLs have advantages when designing the lower risk-adjusted Bernoulli CUSUM chart.

Table 3.5. Estimated in-control performance of lower risk-adjusted Bernoulli CUSUM charts with DPCLs (α = 0.005 and N = 1,000,000)

Risk distribution	Sequence index	$\widehat{\text{ARL}_0}$(S.E.)	$\widehat{\text{SDRL}}$	$Q_{0.10}$	$Q_{0.25}$	$Q_{0.50}$	$Q_{0.75}$	$Q_{0.90}$	$\overline{\alpha_t}$	P_{No}
All	1	385.4(0.97)	308.1	114	169	298	506	790	4.4396E-3	0.2754
	2	393.0(1.04)	327.8	97	152	286	518	824	4.4326E-3	0.2755
	3	380.0(0.99)	314.2	112	154	284	499	803	4.4342E-3	0.2772
	4	379.2(1.00)	316.7	99	156	277	500	790	4.4336E-3	0.2828
	5	377.4(1.00)	315.9	99	149	269	501	797	4.4390E-3	0.2764
High Risk	1	303.7(0.80)	253.4	76	123	230	402	635	4.4544E-3	0.1142
	2	301.6(0.80)	253.2	79	123	222	396	637	4.4604E-3	0.1161
	3	307.1(0.80)	253.1	86	130	222	402	634	4.4582E-3	0.1141
	4	294.5(0.79)	250.1	77	116	215	391	619	4.4582E-3	0.1196
	5	310.0(0.79)	249.7	89	137	230	409	632	4.4549E-3	0.1148
Low Risk	1	422.2(0.77)	242.0	207	247	346	519	740	4.1581E-3	0.0092
	2	405.4(0.74)	233.9	200	240	333	492	708	4.1637E-3	0.0089
	3	411.4(0.76)	241.7	197	237	336	509	729	4.1587E-3	0.0088
	4	409.5(0.75)	237.8	200	238	335	505	717	4.1540E-3	0.0091
	5	417.9(0.75)	238.6	204	245	346	512	731	4.1714E-3	0.0090
Surgeon 1	1	350.3(0.91)	288.4	91	139	264	466	724	4.4226E-3	0.2075
	2	358.3(0.85)	270.2	117	161	271	465	698	4.4193E-3	0.2051
	3	358.7(0.92)	289.5	100	144	264	483	748	4.4311E-3	0.2133
	4	335.0(0.90)	284.1	85	139	251	449	703	4.4241E-3	0.2069
	5	339.9(0.89)	281.5	90	141	252	444	704	4.4212E-3	0.2075
Surgeon 6	1	442.3(1.10)	346.9	138	218	319	563	918	4.4679E-3	0.3457
	2	448.2(1.06)	336.1	158	212	360	573	912	4.4531E-3	0.3341
	3	444.9(1.00)	317.2	161	226	349	559	833	4.4531E-3	0.3332
	4	449.3(1.04)	330.3	153	219	343	572	879	4.4545E-3	0.3350
	5	443.3(1.03)	324.2	150	214	343	569	827	4.4535E-3	0.3372
Geometric	/	200 (/)	200	21	57	138	276	459	/	

Table 3.6. Estimated in-control performance of lower risk-adjusted Bernoulli CUSUM charts with DPCLs (α = 0.001 and N = 1,000,000)

Risk distribution	Sequence index	$\widehat{\mathrm{ARL}_0}$(S.E.)	$\widehat{\mathrm{SDRL}}$	$Q_{0.10}$	$Q_{0.25}$	$Q_{0.50}$	$Q_{0.75}$	$Q_{0.90}$	$\overline{\alpha_t}$	P_{No}
All	1	1225.5(3.50)	1106.6	245	440	886	1639	2670	9.3501E-4	0.0431
	2	1235.5(3.51)	1109.8	249	449	886	1670	2692	9.3442E-4	0.0389
	3	1212.2(3.47)	1098.7	234	429	875	1628	2619	9.3538E-4	0.0401
	4	1236.7(3.54)	1118.0	238	436	887	1665	2676	9.3473E-4	0.0383
	5	1233.5(3.52)	1112.2	242	442	881	1663	2700	9.3608E-4	0.0468
High Risk	1	1120.6(3.32)	1049.7	177	371	800	1530	2485	9.5500E-4	0.0033
	2	1125.1(3.32)	1050.7	190	376	800	1523	2489	9.5621E-4	0.0041
	3	1119.9(3.29)	1041.5	190	379	801	1517	2475	9.5620E-4	0.0043
	4	1142.9(3.32)	1050.6	205	399	821	1547	2505	9.5589E-4	0.0046
	5	1141.5(3.31)	1046.4	203	399	822	1543	2514	9.5562E-4	0.0049
Low Risk	1	1324.5(3.43)	1083.5	358	550	997	1743	2731	9.2670E-4	0.0123
	2	1326.5(3.39)	1071.8	365	562	996	1744	2733	9.2658E-4	0.0124
	3	1326.1(3.40)	1076.4	361	558	998	1745	2726	9.2628E-4	0.0124
	4	1346.7(3.40)	1076.7	384	580	1014	1772	2761	9.2761E-4	0.0134
	5	1322.7(3.39)	1072.3	367	554	987	1748	2728	9.2652E-4	0.0130
Surgeon 1	1	1184.4(3.39)	1072.2	226	412	864	1610	2580	9.4282E-4	0.0167
	2	1167.9(3.35)	1059.1	220	421	845	1574	2538	9.4191E-4	0.0117
	3	1153.9(3.37)	1066.7	197	396	824	1577	2546	9.4126E-4	0.0130
	4	1163.4(3.40)	1074.4	203	401	836	1574	2567	9.4132E-4	0.0134
	5	1179.0(3.40)	1075.7	222	417	847	1586	2580	9.4170E-4	0.0159
Surgeon 6	1	1389.6(3.82)	1209.4	300	540	1018	1871	2992	9.3444E-4	0.1176
	2	1370.4(3.77)	1190.9	307	518	1020	1831	2904	9.3349E-4	0.1142
	3	1415.1(3.83)	1210.5	329	557	1039	1916	3006	9.3394E-4	0.1220
	4	1364.6(3.79)	1199.7	302	495	989	1846	2914	9.3318E-4	0.1173
	5	1386.5(3.70)	1169.5	323	556	1042	1821	2884	9.3263E-4	0.1116
Geometric	/	1000 (/)	1000	105	287	692	1385	2301	/	

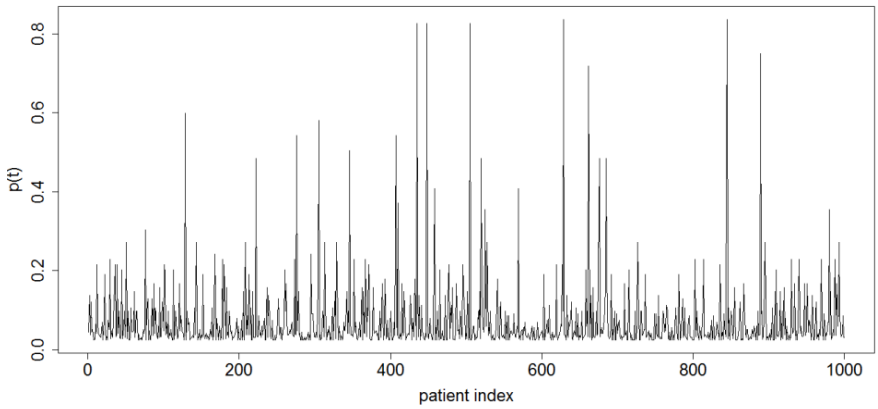
Table 3.7. Estimated in-control performance of lower risk-adjusted Bernoulli CUSUM charts with DPCLs (α = 0.0005 and N = 5,000,000)

Risk distribution	Sequence index	$\widehat{\mathrm{ARL}_0}$(S.E.)	$\widehat{\mathrm{SDRL}}$	$Q_{0.10}$	$Q_{0.25}$	$Q_{0.50}$	$Q_{0.75}$	$Q_{0.90}$	$\overline{\alpha_t}$	P_{No}
All	1	2241.4(6.58)	2079.9	380	756	1609	3054	4951	4.8185E-4	0.00187
High Risk	1	2136.4(6.46)	2042.6	310	686	1513	2917	4795	4.8791E-4	0.00044
Low Risk	1	2335.5(6.53)	2063.8	494	858	1700	3133	5035	4.8236E-4	0.00135
Surgeon 1	1	2177.0(6.54)	2066.9	325	702	1543	2987	4870	4.8404E-4	0.00058
Surgeon 6	1	2348.1(6.81)	2154.5	432	822	1683	3175	5131	4.7970E-4	0.03058
Geometric	/	2000 (/)	2000	210	575	1385	2771	4604	/	/

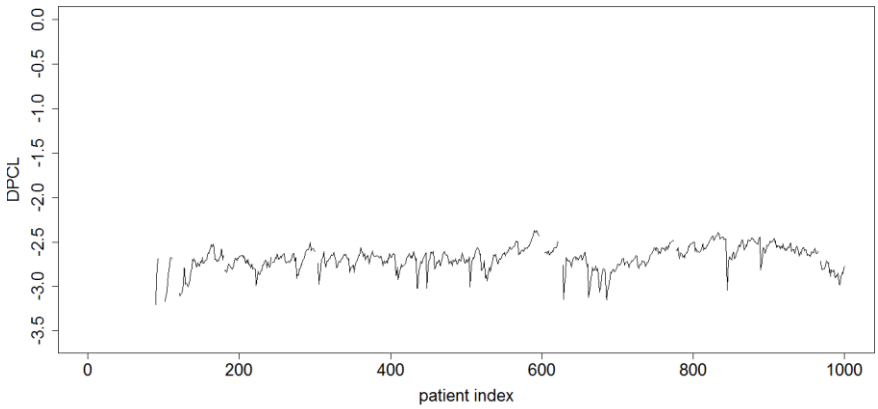
Table 3.8. Estimated in-control performance of lower risk-adjusted Bernoulli CUSUM charts with DPCLs (α = 0.0001 and N = 5,000,000)

Risk distribution	Sequence index	$\widehat{\text{ARL}_0}$(S.E.)	$\widehat{\text{SDRL}}$	$Q_{0.10}$	$Q_{0.25}$	$Q_{0.50}$	$Q_{0.75}$	$Q_{0.90}$	$\overline{\alpha_t}$	P_{No}
All	1	10543.8(32.79)	10369.7	1285	3168	7366	14517	23964	9.6884E-5	0.00093
High Risk	1	10323.9(32.26)	10200.1	1221	3067	7208	14228	23529	9.8073E-5	0.00055
Low Risk	1	10768.3(32.94)	10415.9	1491	3362	7572	14802	24242	9.6103E-5	0.00196
Surgeon 1	1	10384.9(32.26)	10200.4	1222	3101	7254	14404	23665	9.7376E-5	0.00074
Surgeon 6	1	10630.1(32.71)	10344.4	1404	3276	7413	14639	24050	9.6203E-5	0.00157
Geometric	/	10000 (/)	10000	1053	2876	6931	13862	23024	/	/

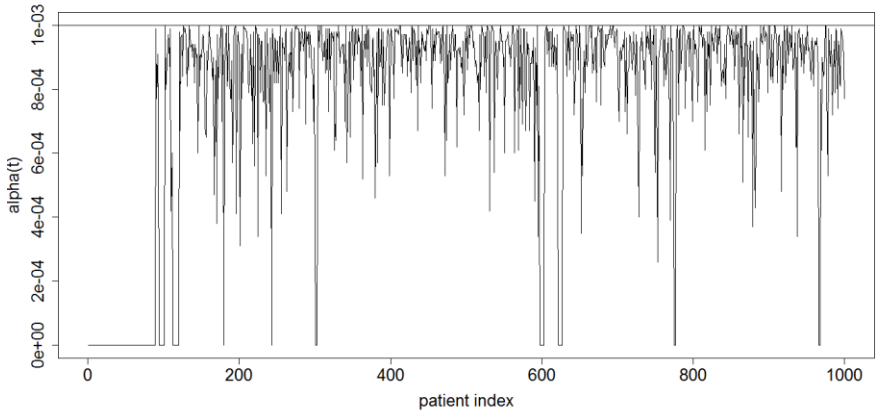
To demonstrate the application of DPCLs to the lower risk-adjusted Bernoulli CUSUM chart, the results for the first 1,000 patients in a sequence from the risk distribution "Surgeon 1" with α = 0.001 are plotted in Figure 3.7. Figure 3.7(a) shows that the surgical failure rates corresponding to the 1,000 patient risk scores vary considerably. Figure 3.7(b) shows the DPCLs obtained using the proposed procedure. There are no control limits for some observations. The reason is that the approximated conditional false alarm rate cannot be set to be less than or equal to α. Thus, there cannot be any signals at those patients for this particular sequence. It can be seen that the probability control limits vary dynamically from patient to patient. Figure 3.7(c) illustrates that the conditional false alarm rate $\alpha_t = \Pr\left\{C_{t,i}^- < h_t^-(0.001)\right\}$ for each patient t is always controlled to be less than or equal to α = 0.001. However, we can see that there are no lower control limits when $\alpha_t = 0$. Therefore, the proportion of no control limits, P_{No}, affects the in-control ARL performance of lower risk-adjusted Bernoulli CUSUM charts with DPCLs.



(a) Estimated surgical failure rate p_t



(b) DPCLs



(c) $\alpha_t = \Pr\left\{C_{t,i}^- < h_t^-(0.001)\right\}$

Figure 3.7. (a) Estimated surgical failure rate p_t, (b) DPCLs, and (c) $\alpha_t = \Pr\left\{C_{t,i}^- < h_t^-(0.001)\right\}$ of the first 1000 patients in a sequence from risk distribution "Surgeon 1" (α = 0.001) for lower risk-adjusted Bernoulli CUSUM chart

4. In-control performance of two-sided risk-adjusted Bernoulli CUSUM charts with DPCLs

In practice, it is common to run both sides of the risk-adjusted Bernoulli CUSUM charts together for detecting deterioration or improvement. Therefore, we apply the DPCLs to the two-sided risk-adjusted Bernoulli CUSUM charts and examine their in-control performance in this section. In theory, to obtain the ARL of the two-sided CUSUM from the ARLs of the two one-sided statistics, say, ARL^+ and ARL^-, the following equation is used,

$$ARL = \frac{ARL^+ \cdot ARL^-}{ARL^+ + ARL^-}, \tag{4}$$

where ARL^+ is calculated for the upper chart and ARL^- the lower chart (Megahed *et al.* (2011)). Figure 3.8 gives an example of DPCLs for a two-sided risk-adjusted Bernoulli CUSUM chart. Here the DPCLs for the first 1000 patients in a sequence from risk distribution "Surgeon 1" (α = 0.001) are used for illustration.

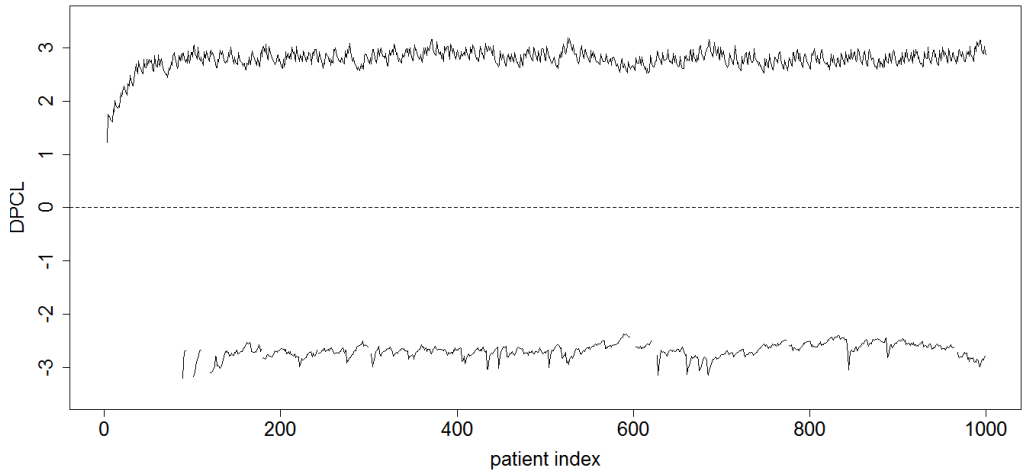


Figure 3.8. An example of DPCLs for a two-sided risk-adjusted Bernoulli CUSUM chart

Next, we simulated 100,000 control charts to estimate the in-control ARL of the two-sided risk-adjusted Bernoulli CUSUM charts with DPCLs for a sequence from each of the five risk distributions with α = 0.005, 0.001, 0.0005, 0.0001. The upper and lower DPCLs applied here had been obtained from previous simulations. From Equation (4), given the estimated upper in-control ARL (ARL_0^+) and the estimated lower in-control ARL (ARL_0^-), the estimated in-control ARL (ARL_0) of two-sided risk-adjusted Bernoulli CUSUM charts with DPCLs should be close to the value $\frac{ARL_0^+ \cdot ARL_0^-}{ARL_0^+ + ARL_0^-}$. Also, the targeted in-control ARL for each false alarm rate $ARL_0 = \frac{ARL_0^+ \cdot ARL_0^-}{ARL_0^+ + ARL_0^-} = \frac{(1/\alpha)^2}{2/\alpha} = (2\alpha)^{-1}$ is given as a reference. The simulation results are shown in Table 3.9.

We note from Table 3.9 that the estimated two-sided in-control ARLs (ARL_0) are all relatively close to the desired values. Therefore, practitioners could run upper and lower risk-adjusted Bernoulli CUSUM charts with DPCLs designed individually and still obtain the desirable in-control performance for the two-sided chart.

Table 3.9. Estimated in-control performance of two-sided risk-adjusted Bernoulli CUSUM charts with DPCLs

α	Risk Distribution	Upper $\widehat{\mathrm{ARL}_0^+}$	Lower $\widehat{\mathrm{ARL}_0^-}$	$\frac{\widehat{\mathrm{ARL}_0^+}\cdot\widehat{\mathrm{ARL}_0^-}}{\widehat{\mathrm{ARL}_0^+}+\widehat{\mathrm{ARL}_0^-}}$	Two-sided $\widehat{\mathrm{ARL}_0}$(S.E.)	Target $\mathrm{ARL}_0 = (2\alpha)^{-1}$
0.005	All	212.7	385.4	137.1	129.6 (0.36)	100
	High Risk	212.6	303.7	125.1	123.6 (0.32)	100
	Low Risk	213.7	422.2	141.9	143.5 (0.36)	100
	Surgeon 1	212.9	350.3	132.4	129.8 (0.34)	100
	Surgeon 6	214.0	442.3	144.2	142.7 (0.35)	100
0.001	All	1027.6	1225.5	558.9	535.7 (1.59)	500
	High Risk	1015.2	1120.6	532.6	515.4 (1.53)	500
	Low Risk	1006.4	1324.5	571.9	577.3 (1.53)	500
	Surgeon 1	1015.4	1184.4	546.7	538.6 (1.55)	500
	Surgeon 6	1018.0	1389.6	587.6	589.6 (1.65)	500
0.0005	All	2018.0	2241.4	1061.9	1048.7 (3.07)	1000
	High Risk	2020.7	2136.4	1038.5	1029.1 (3.27)	1000
	Low Risk	1999.2	2335.5	1077.2	1080.0 (3.02)	1000
	Surgeon 1	2026.3	2177.0	1049.5	1023.8 (3.06)	1000
	Surgeon 6	2006.1	2348.1	1081.8	1072.5 (3.10)	1000
0.0001	All	10027.2	10543.8	5139.5	5038.4 (15.60)	5000
	High Risk	10107.4	10323.9	5107.3	5069.2 (16.93)	5000
	Low Risk	9973.5	10768.3	5177.8	5155.0 (15.83)	5000
	Surgeon 1	10088.5	10384.9	5117.3	5099.2 (16.10)	5000
	Surgeon 6	10049.3	10630.1	5165.8	5117.4 (15.76)	5000

5. Discussion

It is beneficial to obtain evidence of improved surgical performance using the risk-adjusted Bernoulli CUSUM chart. Previous studies showed that the in-control performance of the charts with constant control limits could vary drastically for different risk distributions. In our study, we applied DPCLs to the lower risk-adjusted Bernoulli CUSUM charts and examined the in-control performance of these charts. Although there are deviations from the desired values due to the discreteness of the Bernoulli distribution, the simulation results show that the in-control performance of the charts with DPCLs for different risk distributions varies far less than the chart with fixed control limits. The relationship between in-control performance of the risk-adjusted Bernoulli CUSUM charts with DPCLs and the proportion of patients with no control limits was investigated. In addition, practitioners could run upper and lower risk-adjusted Bernoulli CUSUM charts with DPCLs side by side and obtain the desired in-control performance for the two-sided chart. Finally, our approach could be easily adapted to design the risk-adjusted CUSUM charts based on multiresponses developed by Tang *et al.* (2015) where there can be more than two possible outcomes for each patient.

Acknowledgement

The work of W. H. Woodall was supported by National Science Foundation Grant CMMI-1436365.

References

Beiles, C.B. and Morton, A.P. (2004). "Cumulative Sum Control Charts for Assessing Performance in Arterial Surgery". *ANZ Journal of Surgery,* 74, 3, pp. 146-151.

Bottle, A. and Aylin, P. (2008). "Intelligent Information: A National System for Monitoring Clinical Performance". *Health Services Research,* 43, pp. 10-31.

Chen, T.T., Chung, K.P., Hu, F.C., Fan, C.M., and Yang, M.C. (2011). "The Use of Statistical Process Control (Risk-Adjusted CUSUM, Risk-Adjusted RSPRT and CRAM with Prediction Limits) For Monitoring the Outcomes of Out-of-Hospital Cardiac Arrest Patients Rescued by the EMS System". *Journal of Evaluation in Clinical Practice,* 17, 1, pp. 71-77.

Clinical Practice Improvement Centre. (2008). *VLADs for Dummies.* Wiley Publishing Australia Pty Ltd: Milton, Queensland.

Collins, G.S., Jibawi, A., and McCulloch, P. (2011). "Control Chart Methods for Monitoring Surgical Performance: A Case Study from Gastro-Oesophageal Surgery". *European Journal of Surgical Oncology (EJSO),* 37, 6, pp. 473-480.

Cook, D.A., Duke, G., Hart, G.K., Pilcher, D., and Mullany, D. (2008). "Review of the Application of Risk-Adjusted Charts to Analyze Mortality Outcomes in Critical Care". *Critical Care and Resuscitation: Journal of the Australasian Academy of Critical Care Medicine,* 10, 3, pp. 239-251.

Grigg, O. and Farewell, V. (2004). "An Overview of Risk-Adjusted Charts". *Journal of the Royal Statistical Society: Series A (Statistics in Society),* 167, 3, pp. 523-539.

Harris, J.R., Forbes, T.L., Steiner, S.H., Lawlor, D.K., DeRose, G., and Harris, K.A. (2005). "Risk-Adjusted Analysis of Early Mortality after Ruptured Abdominal Aortic Aneurysm Repair". *Journal of Vascular Surgery*, 42, 3, pp. 387.e1-387.e6.

Hawkins, D.M., Qiu, P., and Kang, C.W. (2003). "The Changepoint Model for Statistical Process Control". *Journal of Quality Technology*, 35, 4, pp. 355-366.

Loke, C.K. and Gan, F.F. (2012). "Joint Monitoring Scheme for Clinical Failures and Predisposed Risks". *Quality Technology and Quantitative Management*, 9, 1, pp. 3-21.

Margavio, T.M., Conerly, M.D., Woodall, W.H., and Drake, L.G. (1995). "Alarm Rates for Quality Control Charts". *Statistics & Probability Letters*, 24, 3, pp. 219-224.

Megahed, F.M., Kensler, J.L., Bedair, K., and Woodall, W.H. (2011). "A Note on the ARL of Two-Sided Bernoulli-Based CUSUM Control Charts". *Journal of Quality Technology*, 43, 1, pp. 43-49.

Moore, R., Nutley, M., Cina, C.S., Motamedi, M., Faris, P., and Abuznadah, W. (2007). "Improved Survival after Introduction of an Emergency Endovascular Therapy Protocol for Ruptured Abdominal Aortic Aneurysms". *Journal of Vascular Surgery*, 45, 3, pp. 443-450.

Morton, A.P., Clements, A.C., Doidge, S.R., Stackelroth, J., Curtis, M., and Whitby, M. (2008). "Surveillance of Healthcare-Acquired Infections in Queensland, Australia: Data and Lessons from the First 5 Years". *Infection Control and Hospital Epidemiology*, 29, 8, pp. 695-701.

Novick, R.J., Fox, S.A., Stitt, L.W., Forbes, T.L., and Steiner, S.H. (2006). "Direct Comparison of Risk-Adjusted and Non–Risk-Adjusted CUSUM Analyses of Coronary Artery Bypass Surgery Outcomes". *The Journal of Thoracic and Cardiovascular Surgery,* 132, 2, pp. 386-391.

Parsonnet, V., Dean, D., and Bernstein, A.D. (1989). "A Method of Uniform Stratification of Risk for Evaluating the Results of Surgery in Acquired Adult Heart Disease". *Circulation,* 79, pp. I3-I12.

Reynolds Jr, M.R. and Stoumbos, Z.G. (1999). "A CUSUM Chart for Monitoring a Proportion When Inspecting Continuously". *Journal of Quality Technology,* 31, 1, pp. 87-108.

Rogers, C.A., Reeves, B.C., Caputo, M., Ganesh, J.S., Bonser, R.S., and Angelini, G.D. (2004). "Control Chart Methods for Monitoring Cardiac Surgical Performance and Their Interpretation". *The Journal of Thoracic and Cardiovascular Surgery,* 128, 1, pp. 811-819.

Shen, X., Zou, C., Jiang, W., and Tsung, F. (2013). "Monitoring Poisson Count Data with Probability Control Limits When Sample Sizes Are Time Varying". *Naval Research Logistics (NRL),* 60, 8, pp. 625-636.

Sherlaw-Johnson, C. (2005). "A Method for Detecting Runs of Good and Bad Clinical Outcomes on Variable Life-Adjusted Display (VLAD) Charts". *Health Care Management Science,* 8, 1, pp. 61-65.

Sherlaw-Johnson, C., Morton, A., Robinson, M.B., and Hall, A. (2005). "Real-Time Monitoring of Coronary Care Mortality: A Comparison and Combination of Two Monitoring Tools". *International Journal of Cardiology,* 100, 2, pp. 301-307.

Sherlaw-Johnson, C., Wilson, P., and Gallivan, S. (2007). "The Development and Use of Tools for Monitoring the Occurrence of Surgical Wound Infections". *Journal of the Operational Research Society,* 58, 2, pp. 228-234.

Steiner, S.H. (2014). "Risk-Adjusted Monitoring of Outcomes in Health Care". Chapter 14 in *Statistics in Action: A Canadian Outlook,* Lawless, J.F., ed., pp. 245–264. Chapman and Hall/CRC.

Steiner, S.H., Cook, R.J., and Farewell, V.T. (2001). "Risk-Adjusted Monitoring of Binary Surgical Outcomes". *Medical Decision Making,* 21, 3, pp. 163-169.

Steiner, S.H., Cook, R.J., Farewell, V.T., and Treasure, T. (2000). "Monitoring Surgical Performance Using Risk-Adjusted Cumulative Sum Charts". *Biostatistics,* 1, 4, pp. 441-452.

Tang, X., Gan, F.F., and Zhang, L. (2015). "Risk-Adjusted Cumulative Sum Charting Procedure Based on Multiresponses". *Journal of the American Statistical Association,* 110, 509, pp. 16-26.

Tian, W., Sun, H., Zhang, X., and Woodall, W.H. (2015). "The Impact of Varying Patient Populations on the In-Control Performance of the Risk-Adjusted CUSUM Chart". *International Journal for Quality in Health Care*, 27, 1, pp. 31-36.

Woodall, W.H. (2006). "The Use of Control Charts in Health-Care and Public-Health Surveillance". *Journal of Quality Technology,* 38, 2, pp. 89-104.

Woodall, W.H., Fogel, S.L., and Steiner, S.H. (2015). "The Monitoring and Improvement of Surgical Outcome Quality". *Journal of Quality Technology*, 47, 4, pp. 383-398.

Zhang, X. and Woodall, W.H. (2015). “Dynamic Probability Control Limits for Risk-Adjusted Bernoulli CUSUM Charts”. Statistics in Medicine, 34, pp. 3336-3348.

Chapter 4

The effect of estimation error on in-control performance of risk-adjusted Bernoulli CUSUM chart with dynamic probability control limits

ABSTRACT

The in-control performance of any control chart is highly associated with the accuracy of estimation for the in-control parameter(s). For the risk-adjusted Bernoulli CUSUM chart with a constant control limit, it had been shown by Jones and Steiner [1] that the estimation error could have a substantial effect on the in-control performance. In our study, we examine the effect of estimation error on the in-control performance of the risk-adjusted Bernoulli CUSUM chart with dynamic probability control limits (DPCLs) developed by Zhang and Woodall [2] using the same data and settings as in the study of Jones and Steiner [1]. The only difference is the way we determine the control limits. Similarly to the study of Jones and Steiner [1], our simulation results show that the in-control performance of risk-adjusted Bernoulli CUSUM chart with DPCLs is also affected by the estimation error. The most important factors affecting estimation error are the specified desired in-control average run length, the Phase I sample size and the adverse event rate. However, the effect of estimation error is uniformly smaller for the risk-adjusted Bernoulli CUSUM chart with DPCLs than for the corresponding chart with a constant control limit under various realistic scenarios. In addition, we found a substantial reduction in the standard deviation of the in-control run length when DPCLs are used. Therefore, use of DPCLs has yet another advantage when designing a risk-adjusted Bernoulli CUSUM chart.

Keywords: average run length (ARL); control limits; estimation error; risk adjustment; statistical process control.

1. Introduction

The design of any control chart involves sampling Phase I data from an assumed stable process and estimating the in-control parameters. These serve as a baseline from which to detect any process change in Phase II. Thus, the control limits are determined based on the estimation of the in-control parameters. The control chart performance will be highly associated with the accuracy of parameter estimation. Jensen *et al.* [3] provided a literature review on the studies assessing effects of parameter estimation on control chart properties in an industrial context. Jones-Farmer *et al.* [4] discussed important aspects of Phase I data collection and analysis.

Driven by the need to appropriately monitor binary surgical outcome quality, Steiner *et al.* [5] developed the now popular risk-adjusted Bernoulli CUSUM chart which takes patients' heterogeneity into account. With this chart one adjusts for each patient's pre-operative risk of surgical failure using a risk-adjustment method, such as a logistic regression model, and then applies the likelihood ratio based scoring method to obtain the monitoring statistics. Here the parameters of in-control risk-adjustment model are estimated using Phase I sample patients. The population of risk factors must also be estimated. Similar to the industrial setting, the performance of risk-adjusted Bernoulli CUSUM chart will be influenced by the accuracy of the parameter estimation. Estimation error will always exist since the population parameters are estimated using sample data which are subject to sampling error due to random variation. Jones and Steiner [1] studied the effect of estimation error on the performance of risk-adjusted CUSUM chart with constant control limits systematically and found that the effect could be substantial. Specifically, the performance was highly dependent on the number of adverse events

used to estimate the control chart parameters and on the specified desired average run length for an in-control process.

Zhang and Woodall [2] recently developed a simulation-based method to determine the dynamic probability control limits (DPCLs) for risk-adjusted Bernoulli CUSUM charts based on the method of Shen *et al.* [6]. The Shen *et al.* [6] method was developed for an application involving the exponentially weighted moving average (EWMA) chart to monitor Poisson count data with time-varying population sizes. By maintaining the probability of a false alarm at a constant level conditional on no false alarm for previous observations, a concept initially proposed by Margavio *et al.* [7], the charts with DPCLs give desirably consistent in-control performance with approximately geometrically distributed in-control run lengths. This method overcomes the disadvantage pointed out by Tian *et al.* [8] of using traditional constant control limits for risk-adjusted Bernoulli CUSUM charts in that the effect of varying risk distributions of patients on the in-control performance is significant [9]. However, in practice, different Phase I data would provide different parameter estimates and therefore different risk-adjustment models. The in-control performance of risk-adjusted Bernoulli CUSUM chart with DPCLs would still be affected by this estimation error, but it will not be affected by changes in the patient risk distribution because the control limits are determined depending on the specific observed sequence of patient risk scores.

In our study, the effect of estimation error on the in-control performance of risk-adjusted Bernoulli CUSUM chart with DPCLs is examined. To make a fair comparison, we use the same data and settings as in the study of Jones and Steiner [1] except for the way to determine the control limits. The control chart performance is usually measured in terms of average run length (ARL) and the standard deviation of the run length (SDRL). Both metrics become random

variables when the in-control parameters are estimated from the Phase I samples. The estimated mean and standard deviation of the ARL are reported as AARL and SDARL, respectively. Similarly, the estimated mean and standard deviation of the SDRL are referred to as ASDRL and SDSDRL. Jones and Steiner [1] originally proposed the SDARL metric with lower values indicating less variation in the ARL values. Due to its simplicity for calculation and interpretation, the SDARL metric has been used for evaluating the performance of various control charts. If the SDARL value is large relative to the desired in-control ARL value, the in-control performance of the control chart would be varying widely for different practitioners with different Phase I samples [10, 11].

The remainder of this paper is organized as follows. In Section 2, the risk-adjusted Bernoulli CUSUM chart and the simulation algorithm to obtain the DPCLs are introduced. Also, the procedure to examine the effect of estimation error on the chart in-control performance is explained. Next, the simulation settings are explained in Section 3. This is followed by the results on how different factors affect the effect of estimation error on in-control performance and a comparison of Jones and Steiner's [1] results with ours in Section 4. Finally, we provide some discussion and related remarks in Section 5.

2. Methods

The one-sided risk-adjusted CUSUM statistics can be written as follows [5],

$$C_t = \max\left(0, C_{t-1} + W_t\right),\ t = 1, 2, 3, \ldots \tag{1}$$

where $C_0 = 0$ and W_t represents the likelihood ratio based weight assigned to the result for the t^{th} individual, which is defined as

$$W_t = \begin{cases} \log\left(\dfrac{1 - p_t + R_0 p_t}{1 - p_t + R_a p_t}\right) & \text{if } y_t = 0, \\ \log\left[\dfrac{\left(1 - p_t + R_0 p_t\right) R_a}{\left(1 - p_t + R_a p_t\right) R_0}\right] & \text{if } y_t = 1. \end{cases} . \tag{2}$$

Here, y_t = 1 if patient *t* experiences the adverse event of interest, such as death or a surgical site infection within a specified time period following surgery, and y_t = 0 otherwise. The risk-adjusted CUSUM chart is designed to monitor for an odds ratio change from R_0 to R_a, where R_0 is usually set to 1 to reflect current expected surgical performance, and set to a value $R_a > R_0$ for detecting process deterioration, and to a value $R_a < R_0$ to detect improvement. The surgical failure rate p_t for each patient is determined by assessing the pre-operative risk of each patient by applying a risk-adjustment method such as a logistic regression model,

$$\text{logit}\left(p_t\right) = \beta_0 + \beta_1 X_t, \tag{3}$$

where X_t is the pre-operative risk score such as Parsonnet score [12] of patient *t*. We let $C_t^+,\ t = 1,\ 2,\ \ldots,$ represent the CUSUM statistics designed to detect an increase in the odds ratio

and C_t^-, $t = 1, 2, \ldots$, the CUSUM statistics designed to detect a decrease in the odds ratio. Following the usual convention we change the signs of the statistics for the lower CUSUM chart and do not allow the values to exceed zero. The chart signals when $C_t^+ > h^+$ or $C_t^- < h^-$ which indicates that there has been either deterioration or improvement in the surgical performance, respectively. The control limits h^+ and h^- are set to yield a suitably large in-control ARL when there are no changes in the odds ratio of failure R_0.

The dynamic probability control limits (DPCLs) are determined by applying the method proposed by Zhang and Woodall [2]. The charts are referred to as risk-adjusted Bernoulli CUSUM charts with DPCLs. For simplicity, we only examine the upper charts for detecting process deterioration (*i.e.* $R_a > R_0$) in our study. The algorithmic form of the simulation procedure to obtain the dynamic probability control limit for patient t (t = 1, 2, 3, …), can be summarized as follows:

1) Generate N Bernoulli random variables $Y_{t,i}$ (i = 1, 2, …, N) with in-control failure rate p_t obtained from risk-adjustment model and N CUSUM statistics $C_{t,i} = \max\left(0, C_{t-1,j} + W_{t,i}\right)$ (j = 1, 2, …, N) using the equations (1) and (2) accordingly, where $C_{t-1,j}$ is randomly selected from a vector of CUSUM values $\boldsymbol{C_{t\text{-}1}}' = \left\{C_{t-1,i} \text{ such that } C_{t-1,i} \le h_{t-1}(\alpha)\right\}$.
2) Sort the N CUSUM statistics in ascending order $C_{t,(1)}$, $C_{t,(2)}$, $C_{t,(3)}$, …, $C_{t,(N)}$ and take the $N' = \lceil N \cdot (1-\alpha) \rceil$ largest CUSUM statistic $C_{t,(N')}$ as the approximated DPCL $h_t(\alpha)$ if $\Pr\left\{C_{t,i} > C_{t,(N')}\right\} \neq 0$.

3) Calculate $C_t^+ = \max\left(0, C_{t-1}^+ + W_t\right)$ based on the surgical outcome and risk score of patient *t*.

If $C_t^+ > h_t(\alpha)$, an out-of-control signal is issued. Otherwise, go back to step 1.

Figure 4.1 shows a comparison of constant control limit and DPCLs for comparable in-control ARLs. The dashed line indicates the constant control limit h^+ = 4.5, while the solid line tracks the DPCLs for a particular sequence of scores chosen from the risk distribution. The advantage of the DPCLs is that one can obtain the desired in-control ARL and maintain the conditional false alarm rates at the specified level without requiring any information or assumptions about the patient population. Also, with DPCLs, it is possible to signal sooner for process changes that occur near the start of the monitoring process [2].

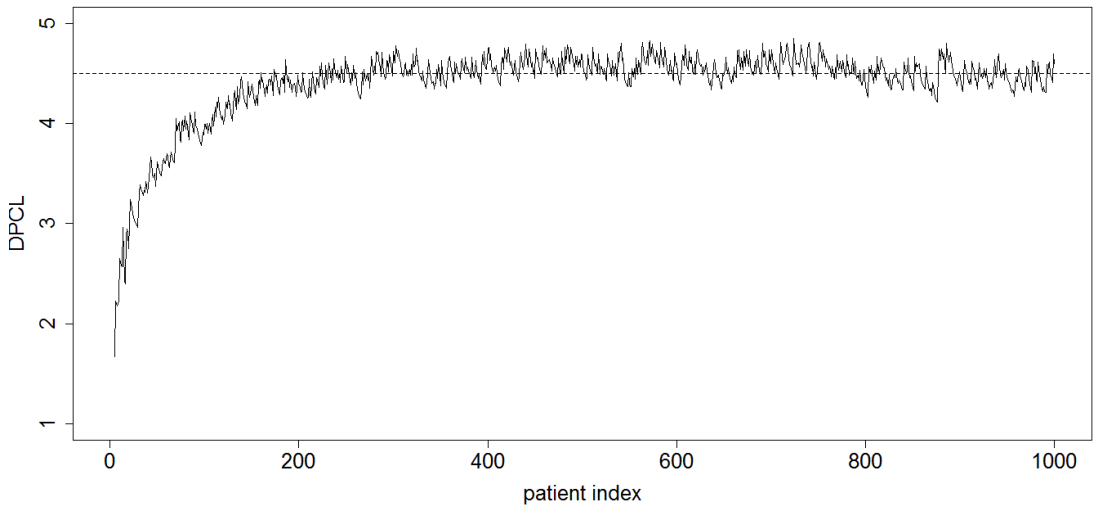


Figure 4.1. Comparison of constant control limit (dashed line) and DPCLs (solid line) for comparable in-control ARLs (α = 0.00013513)

Unlike the risk-adjusted Bernoulli CUSUM charts with constant control limit assessed by Jones and Steiner [1], the ones with DPCLs have been shown to be able to deliver consistent desirable in-control performance regardless of the patient population [2]. Therefore, one does not

have to estimate the patient mix as with their approach. Moreover, it can be clearly seen from equations (2) and (3), that the risk-adjusted CUSUM weight W_t is determined by the estimated surgical failure rate p_t, which is calculated from the risk-adjustment logistic regression model. As a result, the in-control performance of the chart would be associated with accuracy of the estimation for parameters β_0 and β_1 from the sample data. By assuming that there is no confounding bias and other important covariates are not missing in the risk-adjustment model, Phase I data are sampled from the patient population and the logistic regression model is estimated. Also we assume that there are no errors in the measurement of the covariates (*i.e.* the risk scores) and the covariates are correctly used in building the risk-adjustment model.

To examine the effect of estimation error on the in-control performance of the chart, the following steps were repeated,

1) Sample Phase I data (X_t, y_t), $t = 1, 2, \ldots, n$, from the assumed true patient population.
2) Use the data to estimate the parameters of the risk-adjustment model, β_0 and β_1.
3) Use the sample data and risk-adjustment model to determine the dynamic probability control limits (DPCLs) for a sequence of patients randomly sampled with replacement from the Phase I sample data. The conditional false alarm rate is set to be α in order to yield the specified desired value of in-control ARL as $1/\alpha$. The method of determining the DPCLs is simulation-based.
4) Given the obtained DPCLs and the true risk-adjustment model, estimate the actual in-control ARL (ARL_0) for the sequence of patients selected in step 3 by simulation.

The distribution of the actual ARL_0 is obtained by repeating the steps 1 – 4 a considerable number of times, say 1000 times. The effect of estimation error is represented by the variation

among the ARL_0's. The average ARL_0 ($AARL_0$) and standard deviation of ARL_0 ($SDARL_0$) are reported in the results section for each setting explained in the next section. Moreover, the average standard deviation of in-control run lengths ($ASDRL_0$) and the standard deviation of $SDRL_0$ ($SDSDRL_0$) are also listed for each setting.

3. Simulation settings

To have a fair comparison, we used the same data and approach as in the study of Jones and Steiner [1]. Both actual and simulated data were used to assess the effect of estimation error in their study. The actual data we used were the same used by Steiner *et al.* [5] on a cohort of 6994 patients from the UK who underwent coronary bypass graft surgery (CABG) between Jan 1st, 1992 and Dec 31st, 1998. The 30-day mortality was considered as the adverse event of primary interest which occurred in 6.6% of the patients. It was assumed that the patient mix of the 6994 patients and the fitted logistic model, $\text{logit}(p_t) = -3.63 + 0.074X_t$, that described the relationship between Parsonnet score and 30-day mortality rate were correct so that this could serve as a baseline. We note that only the 2218 patients from the first two years were treated as in-control group and used to fit the logistic regression risk model in some previous related studies [2, 5, 8]. However, here we used all 6994 patients as in-control group. Also, it was assumed that a doubling of the odds of mortality is a clinically meaningful increase that is to be detected quickly.

First, we examined how different sizes of Phase I data impacted the effect of estimation error. The sample sizes were set as 760, 1520, 2280, 3040 and 3800 patients so that the expected number of 30-day deaths equal 50, 100, 150, 200 and 250, respectively. One thousand samples of each sample size were randomly drawn with replacement from the 6994 patients. The conditional false alarm rate α was set to be 0.001 resulting in the desired in-control ARL of approximately 1000.

Second, for the same sample size, we investigated how different conditional false alarm rates influenced the effect of estimation error. Three values of α, 1/500, 1/1000 and 1/1500, were examined and 1000 samples of 1520 patients for each of the three scenarios were taken.

Third, the adverse event rate was changed to 3.5% and also to 12.2% by modifying the logistic regression model to $\text{logit}(p_t) = -4.36 + 0.074X_t$ and $\text{logit}(p_t) = -2.88 + 0.074X_t$, which were considered as "true" models, respectively. For each of these two models, we sampled 1520 Parsonnet scores from the CABG data set with replacement and estimated p_t for each score X_t. A Bernoulli random outcome z_t with $P(\text{adverse event}) = p_t$ was generated for each score X_t. Then we used (X_t, z_t), $t = 1, 2, \ldots, 1520$, to fit a logistic regression model for step 2 in the previous section. The same procedure was conducted 1000 times for each modified model.

Finally, we only drew samples from the patients with a Parsonnet score of 20 or less. The model for these lower-risk patients then became $\text{logit}(p_t) = -4.02 + 0.114X_t$. Also we assumed another scenario with completely homogeneous patients where the adverse event risk was the same for each patient. The model was $\text{logit}(p_t) = -2.65$ resulting in a uniform risk of 6.6% for each patient. For each model, we sampled 1520 Parsonnet scores from the CABG data set with replacement and estimated p_t for each score X_t. A Bernoulli random outcome z_t with $P(\text{adverse event}) = p_t$ was generated for each score X_t. Then a logistic regression model for step 2 in the previous section was fitted using (X_t, z_t), $t = 1, 2, \ldots, 1520$. We conducted the same procedure 1000 times for each modified model.

4. Results

The simulation results are shown in Table 4.1 through Table 4.4. Both results from Jones and Steiner [1] and from our study for each simulation setting are listed and compared in each table. Table 4.1 shows the effect of different Phase I sample sizes on estimation error based on 1000 samples chosen for each of the five sample sizes. The $SDARL_0$ and $SDSDRL_0$ of risk-adjusted Bernoulli CUSUM (RA-CUSUM) charts with DPCLs decrease as the sample size increases. More specifically, both $SDARL_0$ and $SDSDRL_0$ decrease by a factor of $\sqrt{2}$ as the sample size doubles. This finding is similar to the study of Jones and Steiner [1]. However, it can be clearly seen that the $SDARL_0$, $ASDRL_0$ and $SDSDRL_0$ of the RA-CUSUM charts with DPCLs are uniformly smaller compared to the RA-CUSUM charts with constant control limits. Since one can obtain approximately geometrically distributed run lengths by applying DPCLs to the RA-CUSUM chart, it is unsurprising to observe that $ASDRL_0$ values are very close to the $AARL_0$ values for all the sample sizes in our study. The $ASDRL_0$ values are about 40% larger than the $AARL_0$ values if constant control limits are used.

Table 4.1. Estimated in-control performance for each of the five Phase I sample sizes

Sample size	Jones and Steiner (2012)			Performance of RA-CUSUM with DPCLs		
	Average number of deaths (SD)	$AARL_0$ ($SDARL_0$)	$ASDRL_0$ ($SDSDRL_0$)	Average number of deaths (SD)	$AARL_0$ ($SDARL_0$)	$ASDRL_0$ ($SDSDRL_0$)
760	50.1 (7.2)	1008 (131.5)	1398 (184)	50.0 (6.6)	999 (117.6)	997 (113.3)
1520[a]	100.0 (9.6)	1011 (89.2)	1402 (125)	100.0 (9.6)	1005 (85.0)	1001 (83.7)
2280	150.8 (12.2)	1016 (73.6)	1408 (103)	149.7 (11.4)	1002 (64.0)	999 (64.0)
3040	200.8 (13.7)	1015 (64.4)	1408 (90)	201.2 (13.6)	1011 (57.5)	1007 (57.1)
3800	249.9 (14.8)	1014 (56.8)	1407 (79)	250.5 (15.2)	1007 (52.4)	1004 (52.7)

[a]This is the same reference scenario in each of the four tables.

Table 4.2 shows the effect of different conditional false alarm rates (α) on the effect of estimation error. In other words, we examine how specifying a different desired in-control ARL impacts the effect of estimation error. As the specified value doubles, both $SDARL_0$ and $SDSDRL_0$ double. This relationship was also shown in Jones and Steiner's study. Similarly to Table 4.1, the $SDARL_0$, $ASDRL_0$ and $SDSDRL_0$ of RA-CUSUM with DPCLs are uniformly smaller under all scenarios compared to the results in Jones and Steiner's study [1] with the use of the constant control limit. These lower values indicate less variation in the performance of the chart with DPCLs.

Table 4.2. Estimated in-control performance for each of the three specified in-control ARL values

Specified in-control ARL (α)	Jones and Steiner (2012)			Performance of RA-CUSUM with DPCLs		
	Average number of deaths (SD)	$AARL_0$ ($SDARL_0$)	$ASDRL_0$ ($SDSDRL_0$)	Average number of deaths (SD)	$AARL_0$ ($SDARL_0$)	$ASDRL_0$ ($SDSDRL_0$)
500 (1/500)	100.4 (9.5)	504 (43.2)	695 (60)	100.1 (9.7)	504 (40.1)	503 (40.3)
1000[a] (1/1000)	100.0 (9.6)	1011 (89.2)	1402 (125)	100.0 (9.6)	1005 (85.0)	1001 (83.7)
1500 (1/1500)	100.6 (9.5)	1526 (129.9)	2123 (182)	100.1 (9.8)	1519 (125.7)	1516 (122.6)

[a]This is the same reference scenario in each of the four tables.

The influence of different adverse event rates on the effect of estimation error is shown in Table 4.3. It can be seen that there is a relationship between adverse event rate and estimation error similar to the one between sample size and estimation error. As the adverse event rate (approximately) doubles, both $SDARL_0$ and $SDSDRL_0$ decrease by a factor of $\sqrt{2}$. Compared to Jones and Steiner's [1] study, we still observe the uniformly smaller $SDARL_0$, $ASDRL_0$ and $SDSDRL_0$ values in the performance of RA-CUSUM with DPCLs. The reason for the increased

variation in the run lengths and the average run lengths for the RA-CUSUM chart with a constant control limit is that the run length distribution and average run length will vary for a specific risk population depending on which sequence of patients is observed. This is not a factor for the RA-CUSUM chart with DPCLs because the limits are determined for each particular sequence of patients.

Table 4.3. Estimated in-control performance for each of the three overall adverse event rates

Overall adverse event rate	Jones and Steiner (2012)			Performance of RA-CUSUM with DPCLs		
	Average number of deaths (SD)	$AARL_0$ ($SDARL_0$)	$ASDRL_0$ ($SDSDRL_0$)	Average number of deaths (SD)	$AARL_0$ ($SDARL_0$)	$ASDRL_0$ ($SDSDRL_0$)
3.5%	53.8 (7.5)	989 (125.9)	1367 (175)	54.1 (7.2)	1011 (110.6)	1016 (111.9)
6.6%[a]	100.0 (9.6)	1011 (89.2)	1402 (125)	100.0 (9.6)	1005 (85.0)	1001 (83.7)
12.2%	185.2 (12.9)	1001 (60.5)	1392 (85)	184.9 (12.2)	1004 (56.7)	1002 (56.4)

[a]This is the same reference scenario in each of the four tables.

We show the effect of different levels of patient variation on the effect of estimation error in Table 4.4. Here the scenario "Low" only considers the patients with a Parsonnet score of 20 or less which have the overall adverse event rate of 4.6%. And "High" represents the overall patient population which includes not only lower risk patients but also the ones with a Parsonnet score of higher than 20. The results demonstrate that the RA-CUSUM chart with DPCLs outperforms the one with constant control limits in the sense that it has smaller $SDARL_0$, $ASDRL_0$ and $SDSDRL_0$ values. Similar to Jones and Steiner's study [1], however, the patient variability has only a modest effect on the effect of estimation error for the RA-CUSUM chart with DPCLs. It can be seen that the difference in the effect of estimation error is actually associated with the difference in the adverse event rates among the patient populations for the RA-CUSUM chart

with DPCLs. If the overall adverse event rate remains the same, the effect of estimation error on the in-control performance of RA-CUSUM chart with DPCLs stays approximately the same regardless of the level of patient variation.

Table 4.4. Estimated in-control performance for each of three levels of patient variation

Level of patient variation	Jones and Steiner (2012)			Performance of RA-CUSUM with DPCLs		
	Average number of deaths (SD)	$AARL_0$ ($SDARL_0$)	$ASDRL_0$ ($SDSDRL_0$)	Average number of deaths (SD)	$AARL_0$ ($SDARL_0$)	$ASDRL_0$ ($SDSDRL_0$)
High[a]	100.0 (9.6)	1011 (89.2)	1402 (125)	100.0 (9.6)	1005 (85.0)	1001 (83.7)
Low	69.5 (7.9)	1003 (103.4)	1390 (144)	70.5 (8.2)	1014 (97.4)	1013 (96.1)
None	100.4* (9.5*)	1000* (85.9*)	1389* (120*)	100.3 (9.5)	1000 (85.0)	999 (84.0)

[a]This is the same reference scenario in each of the four tables.
*Updated values from MA Jones, personal communication.

5. Discussion

In our study, we demonstrate that the in-control performance of risk-adjusted Bernoulli CUSUM chart with dynamic probability control limits is less affected by estimation error than the RA-CUSUM chart with a constant control limit. The most important factors affecting estimation error are the specified desired in-control average run length, the Phase I sample size and the overall adverse event rate. First, as the specified in-control ARL doubles, both $SDARL_0$ and $SDSDRL_0$ double approximately. Second, both $SDARL_0$ and $SDSDRL_0$ decrease by a factor of $\sqrt{2}$ approximately as the Phase I sample size doubles or the adverse event rate doubles. These finding are similar to the study of Jones and Steiner [1] on the effect of estimation error on the RA-CUSUM chart with a constant control limit. Although the $AARL_0$ can be controlled at the desired level, the variation among ARL_0's due to estimation error is not negligible. Therefore, updating the in-control risk-adjustment model was recommended. We note that our study was set up in the same way as Jones and Steiner's study [1] except we used DPCLs and simulation to estimate the in-control ARL (ARL_0) instead of using a Markov chain approximation.

We found that the effect of estimation error is uniformly smaller for the RA-CUSUM chart with DPCLs than the chart with a constant control limit under various realistic scenarios. Since one can obtain approximately geometrically distributed run lengths by applying DPCLs to the RA-CUSUM chart, the $ASDRL_0$ values are very close to the $AARL_0$ values for all the scenarios in our study. However, the $ASDRL_0$ values are about 40% larger than the $AARL_0$ values if constant control limits are applied with the RA-CUSUM chart. The better performance is due to the fact that applying DPCLs does not require the estimation of the patient mix. Thus,

use of DPCLs is demonstrated to have another advantage when designing a risk-adjusted Bernoulli CUSUM chart.

Ways to reduce the effect of estimation error on the in-control performance of the RA-CUSUM chart with DPCLs should be the future direction of studies. For instance, using a bootstrap method [1] to determine the estimated surgical failure rates and the control limits could be considered.

References

1. Jones MA, Steiner SH. Assessing the effect of estimation error on risk-adjusted CUSUM chart performance. *Int J Qual Health Care* 2012;24:176-81.

2. Zhang X, Woodall WH. Dynamic probability control limits for risk-adjusted Bernoulli CUSUM charts. *Stat Med* 2015;34:3336–48.

3. Jensen WA, Jones-Farmer LA, Champ CW, Woodall WH. Effects of parameter estimation on control chart properties: a literature review. *J Qual Technol* 2006;38:349-64.

4. Jones-Farmer LA, Woodall WH, Steiner SH, Champ CW. An overview of phase I analysis for process improvement and monitoring. *J Qual Technol* 2014;46:265-80.

5. Steiner SH, Cook RJ, Farewell VT, Treasure T. Monitoring surgical performance using risk-adjusted cumulative sum charts. *Biostatistics* 2000;1:441-52.

6. Shen X, Zou C, Jiang W, Tsung F. Monitoring Poisson count data with probability control limits when sample sizes are time varying. *Nav Res Log* 2013;60:625-36.

7. Margavio TM, Conerly MD, Woodall WH, Drake LG. Alarm rates for quality control charts. *Stat Probabil Lett* 1995;24:219-24.

8. Tian W, Sun H, Zhang X, Woodall WH. The impact of varying patient populations on the in-control performance of the risk-adjusted CUSUM chart. *Int J Qual Health Care* 2015;27:31-6.

9. Woodall WH, Fogel SL, Steiner SH. The Monitoring and Improvement of Surgical Outcome Quality. *J Qual Technol* 2015;47:383-98.

10. Saleh NA, Mahmoud MA, Keefe MJ, Woodall WH. The difficulty in designing Shewhart $\bar{X}$ and *X* control charts with estimated parameters. *J Qual Technol* 2015;47:127-38.

11. Saleh NA, Mahmoud MA, Jones-Farmer LA, Zwetsloot I, Woodall WH. Another look at the EWMA control chart with estimated parameters. *J Qual Technol* 2015;47:363-82.

12. Parsonnet V, Dean D, Bernstein AD. A method of uniform stratification of risk for evaluating the results of surgery in acquired adult heart disease. *Circulation* 1989;79:I3-12.

Chapter 5

Dynamic probability control limits for risk-adjusted CUSUM charts based on multiresponses

ABSTRACT

For a patient who has survived a surgery, there could be several levels of recovery. Thus, it is reasonable to consider more than two outcomes when monitoring surgical outcome quality. Tang *et al.* (2015) developed the risk-adjusted cumulative sum (CUSUM) chart based on multiresponses for monitoring a surgical process with three or more outcomes. Similar to the risk-adjusted Bernoulli CUSUM chart proposed by Steiner *et al.* (2000), however, there is a significant effect of varying risk distributions on the in-control performance of the risk-adjusted CUSUM chart for multiresponses when constant control limits are applied. In our study, it is shown that the in-control average run lengths (ARLs) of risk-adjusted CUSUM charts for multiresponses with constant control limits and a given risk-adjustment model can vary by a factor of two for different realistic patient populations. To overcome this disadvantage, we apply the dynamic probability control limits (DPCLs) proposed by Zhang and Woodall (2015). The simulation results demonstrate that the in-control performance of the charts with DPCLs can be controlled for different patient populations because these limits are determined for each specific sequence of patients. Thus, the use of DPCLs for risk-adjusted CUSUM charts based on multiresponses allows each chart to be designed for the corresponding patient sequence of a surgeon or a hospital and therefore does not require estimating or monitoring of the patients' risk distribution.

Keywords: average run length (ARL); false alarm rate; proportional odds logistic regression; statistical process monitoring; surgical performance.

1. Introduction

Healthcare quality has been an important topic of discussion for many years. Recently, increasing attention has been placed upon monitoring the quality of healthcare performance. The use of statistical process control techniques to appropriately monitor the quality of healthcare has garnered increasing interest due to its impact and effectiveness in other fields such as manufacturing and public health surveillance. One characteristic of interest which has received significant attention is that of surgical outcome quality. The timely detection of any deterioration in surgical performance is the key to prompt investigations of possible causes and to avoid adverse consequences. Moreover, it is quite beneficial for healthcare practitioners and organizations to obtain evidence of improved surgical performance. Woodall *et al.* (2015) gave a thorough review on monitoring and improving surgical outcome quality.

Unlike the process monitoring in industrial quality control applications where the items being produced are typically assumed to be relatively homogeneous, monitoring the quality of surgical outcome performance has to take into account the heterogeneity in patients. To make meaningful comparisons between surgeons or hospitals or to adequately monitor the surgical outcome quality, a risk-adjustment model is needed to account for the variability among surgical patients with respect to their physical characteristics and health status. Grigg and Farewell (2004), Woodall (2006), Cook *et al.* (2008), Steiner (2014) and Woodall *et al.* (2015) provided overviews of risk-adjusted monitoring.

Steiner *et al.* (2000) and Steiner *et al.* (2001) developed the risk-adjusted Bernoulli CUSUM chart which involves adjusting for each patient's pre-operative risk of surgical failure using a logistic regression model and then applying a likelihood-ratio based scoring method to

obtain the monitoring statistics. Their risk-adjusted Bernoulli CUSUM chart has been shown to be well suited for settings where there is a mix of patients with various pre-operative risks and therefore has been recommended and applied by a number of practitioners for various applications. The risk-adjusted CUSUM charts has been used to assess or monitor clinical outcome performance for various applications (Beiles *et al.* 2004; Harris *et al.* 2005; Sherlaw-Johnson 2005; Sherlaw-Johnson *et al.* 2005; Novick *et al.* 2006; Moore *et al.* 2007; Sherlaw-Johnson *et al.* 2007; Bottle and Aylin 2008; Morton *et al.* 2008; Collins *et al.* 2011; and Chen *et al.* 2011).

An obvious drawback to the risk-adjusted Bernoulli CUSUM chart, however, is the binary restriction placed on the outcome of interest. As Tang *et al.* (2015) argued, it is not reasonable to classify a patient who has a full recovery and a patient who has survived but remained in poor condition into the same surgical outcome category. For a patient who survives a surgery, there can be several levels of recovery. Lovegorove *et al.* (1997) and Shortell *et al.* (2000) provided classifications considering more than two outcomes. Ryan *et al.* (2011) mentioned the idea of monitoring with more than two outcomes. Tang *et al.* (2015) also pointed out that "the outcomes from a surgery can be more meaningfully and usefully classified into more than two outcomes". To incorporate more useful information, Tang *et al.* (2015) developed a risk-adjusted CUSUM charting procedure based on more than two outcomes which uses a proportional odds logistic regression model to estimate the probabilities of various surgical outcomes. It was demonstrated that the risk-adjusted CUSUM chart for multiresponses is more sensitive in detecting deterioration of surgical performance in the presence of a substantial proportion of partial recoveries than the risk-adjusted Bernoulli CUSUM chart.

In the design of any CUSUM procedure, it is important to set appropriate control limits to obtain a desired in-control average run length (ARL_0) value. The run length is defined in surgical performance monitoring as the number of trials, i.e., patients, observed until a signal is given by the control chart indicating that a process change may have occurred. An ideal chart is one with a relatively large ARL_0 value when there are no changes in the parameter of interest, whereas a small out-of-control average run length (ARL_1) value is desired when the parameter of interest has actually changed. Moreover, getting the same or very close ARL_0 values is the prerequisite for comparing the out-of-control performance of competing control charts. Tang *et al.* (2015) did not determine the control limits for their method based on the desired value of ARL_0. Instead they somewhat arbitrarily set the control limit to be 3.5 times the standard deviation of the steady-state CUSUM statistic under the baseline model.

In the cardiac surgery examples shown in Steiner *et al.* (2000), the control limits for the proposed CUSUM charts were set at specified levels to give the relatively large ARL_0 values given the patient population and the fitted risk-adjustment models. However, several researchers have brought up issues about the effect of different risk distributions on the performance of risk-adjusted CUSUM charts with constant control limits. See Steiner *et al.* (2001), Rogers *et al.* (2004), Loke and Gan (2012) and Tian *et al.* (2015). The ARL_0 values of risk-adjusted CUSUM charts with the same risk adjustment model and constant control limits can vary by a factor of ten for the highest and lowest risk patient populations. Tang *et al.* (2015) stated that "… the underlying risk distribution of the patients has an effect on the performance of a CUSUM chart." Different surgeons or hospitals could have quite different mixes of patients. Therefore, use of constant control limits for the risk-adjusted CUSUM charts could lead to quite variable in-control performance.

To overcome this disadvantage of using constant control limits, Zhang and Woodall (2015) recently developed a simulation-based procedure to determine the dynamic probability control limits (DPCLs) for risk-adjusted Bernoulli CUSUM charts based on the method of Shen *et al.* (2013). The concept in non-dynamic applications was previously used by Margavio *et al.* (1995) and Hawkins (2003). By maintaining the conditional false alarm rate at a constant value given that there are no false alarms for previous observations, one can design the risk-adjusted CUSUM chart with DPCLs which result in approximately geometrically distributed in-control run lengths with a desired in-control ARL for any sequence of patients. The simulation results of applying DPCLs to upper risk-adjusted Bernoulli CUSUM charts showed that the method does not rely on any assumptions about the patients' risk distribution. Therefore, using DPCLs instead of constant control limits overcomes the major disadvantage of the risk-adjusted CUSUM charts and is more practical for monitoring surgical outcome quality. One no longer has to estimate or monitor the risk distribution as recommended by Steiner *et al.* (2000) and Tang *et al.* (2015).

In our study, we apply the proposed DPCLs to the risk-adjusted CUSUM chart for multiresponses. The remainder of this paper is organized as follows. The risk-adjusted CUSUM chart for multiresponses is introduced in Section 2 with an examination of the effect of varying patient populations on the in-control performance when constant control limits are applied. Next, the procedure to determine the DPCLs for risk-adjusted CUSUM charts based on multirepsonses is explained in Section 3. This is followed by a discussion in Section 4 of the simulation study, including simulation settings and results demonstrating the desirable in-control performance obtained by applying DPCLs. Finally, some discussion and conclusions are provided in Section 5.

2. Risk-adjusted CUSUM chart for multiresponses

2.1. Method

The basic procedure to design the risk-adjusted CUSUM chart for multiresponses developed by Tang *et al.* (2015) is introduced in this section.

Generally, the one-sided tabular CUSUM statistics (Page (1954)) can be written as

$$C_t = \max\left(0, C_{t-1} + W_t\right),\ t = 1, 2, 3, \ldots \tag{1}$$

where $C_0 = 0$ and W_t represents the weight assigned to the result for the t^{th} individual. To monitor a specified change of the parameter of interest θ, say from θ_0 to θ_a, the optimal choice for W_t is the log-likelihood ratio $W_t = \ln\left(f\left(y_t;\theta_a\right)/f\left(y_t;\theta_0\right)\right)$, where y_t is the outcome for the t^{th} individual. The chart signals when C_t exceeds the specified control limit *h*.

For monitoring surgical outcome quality, the risk score such as Parsonnet score (Parsonnet *et al.* (1989)) S_t for each patient *t* is measured and taken into account to estimate the surgical failure rate. In the approach of Tang *et al.* (2015), the outcome of a surgery can be represented by an integer random variable *Y* which takes a value from 0 to *J*, where $Y = J$ when the patient dies within 30 days after surgery, $Y = 0$ when the patient has a full recovery, and $Y = 1, 2, \ldots, J-1$ represent different mutually exclusive ordered states of partial recovery, with a smaller number corresponded to a better state of recovery. Conditional on the patient's risk score $S = s$, we have

$$P\left(Y=k\left|S=s\right.\right)=\pi_k(s),\ k=0,\ 1,\ \ldots,\ J\ \text{ and }\ \sum_{k=0}^{J}\pi_k(s)=1.$$

The multiple response logistic regression model is used to estimate the probability mass function of Y using a historical data set of patients' risk scores and their surgical outcomes. The cumulative logit is defined as

$$\text{logit}\left[P\left(Y\le k\left|S=s\right.\right)\right]=\log\left[\frac{P\left(Y\le k\left|S=s\right.\right)}{1-P\left(Y\le k\left|S=s\right.\right)}\right]=\log\left[\frac{\pi_0(s)+\cdots+\pi_k(s)}{\pi_{k+1}(s)+\cdots+\pi_J(s)}\right],\qquad(2)$$

where $k=0,\ 1,\ \ldots,\ J-1$. The proportional odds logistic regression model (McCullagh, 1980) uses all the J cumulative logits simultaneously as follows,

$$\text{logit}\left[P\left(Y\le k\left|S=s\right.\right)\right]=\alpha_k+\beta s,\ k=0,\ 1,\ \ldots,\ J-1.\qquad(3)$$

Thus, each cumulative logit is assumed to have its own intercept α_k and share the same slope β. Under the in-control and out-of-control situations, we assume $\left(\pi_0(s),\ \ldots,\ \pi_J(s)\right)$ = $\left(\pi_0^0(s),\ \ldots,\ \pi_J^0(s)\right)$ and $\left(\pi_0(s),\ \ldots,\ \pi_J(s)\right)=\left(\pi_0^A(s),\ \ldots,\ \pi_J^A(s)\right)$ respectively. Tang *et al*. (2015) showed that the risk-adjusted CUSUM weight W_t can be written as

$$W_t=\ln\left(\pi_{Y_t}^A(S_t)\big/\pi_{Y_t}^0(S_t)\right).\qquad(4)$$

For the multiresponses proportional odds logistic regression model, the performance is defined based on cumulative probabilities,

$$\frac{\sum_{i=0}^{k}\pi_i^*(s)}{1-\sum_{i=0}^{k}\pi_i^*(s)} = R_k \frac{\sum_{i=0}^{k}\pi_i(s)}{1-\sum_{i=0}^{k}\pi_i(s)}, \quad k = 0,\ 1,\ \ldots,\ J-1, \tag{5}$$

where R_k is the odds ratio of cumulative probabilities of different stage of recovery. In practice, it is common to assume $R_0 = R_1 = \ldots = R_{J-1} = 1$ when the process is in-control. One sets $R_0 = R_1 = \ldots = R_{J-1} > 1$ for detecting improvement and $R_0 = R_1 = \ldots = R_{J-1} < 1$ for detecting deterioration.

The procedure of Tang *et al*. (2015) for constructing the risk-adjusted CUSUM chart based on multiple responses can be summarized as follows:

1) Fit a proportional odds logistic regression model (3) using historical Phase I data.
2) Estimate the probability of obtaining the outcome *k*, $\pi_k(s)$, $k = 0,\ 1,\ \ldots,\ J$, for a patient with risk score *s* using the fitted logistic regression model to obtain the in-control probability.
3) Set the out-of-control values $R_0 = R_1 = \ldots = R_{J-1} = R_A$, then use (5) to estimate the probability of obtaining the outcome *k*, $\pi_k^*(s)$, $k = 0,\ 1,\ \ldots,\ J$, for a patient with risk score *s* under the process shift of interest.
4) Calculate the CUSUM weight $W_t = \ln\left(\pi_{Y_t}^A(S_t)/\pi_{Y_t}^0(S_t)\right) = \ln\left(\pi_k^*(s)/\pi_k(s)\right)$ based on the surgical outcome $Y_t = k$ and risk score $S_t = s$ of patient *t*.
5) Calculate CUSUM statistics $C_t = \max\left(0, C_{t-1} + W_t\right)$, t = 1, 2, 3, …, where $C_0 = 0$. If C_t exceeds the control limit which is set to yield a desired in-control average run length

(ARL_0), an out-of-control signal is issued. Otherwise, go back to step 4 with the value of t increased by 1.

2.2. Effect of varying patient populations on the in-control performance

Tang *et al.* (2015) mentioned in their paper that the underlying risk distribution of the patients has an effect on the performance of their CUSUM chart. Recently, Tian *et al.* (2015) examined the effect of varying patient populations on the in-control performance of the risk-adjusted Bernoulli CUSUM chart. The simulation study showed that the in-control ARLs of risk-adjusted Bernoulli CUSUM charts with fixed control limits and a given risk-adjustment model can vary by a factor of two for different realistic patient populations. Varying patient populations are very common in applications. It is realistic to assume that the patient mixes for different hospitals and different surgeons can vary considerably, as well as vary over time. In this section, we examine closely the effect of varying patient populations on the in-control performance of risk-adjusted CUSUM charts for multiresponses with constant control limits.

We used the same data set from a seven-year study used by both Steiner *et al.* (2000) and Tang *et al.* (2015). Tang *et al.* (2015) considered the surgical outcome $Y = 2$ if the patient died within 30 days of surgery, $Y = 1$ if the patient died after 30 days, but before the end of the seven-year study, and $Y = 0$ if the patient was still surviving when the data were finalized. This classification system, however, would be unrealistic for use in practice since it could not be employed prospectively. One would have to wait until the end of the study to plot statistics for the patients with outcome $Y = 0$ or 1. In prospective monitoring, data collection is assumed to be ongoing with no planned end of the study. Therefore, we use a different classification in our study with $Y = 2$ if the patient died within 14 days, $Y = 1$ if the patient died after 14 days but

within 30 days and $Y = 0$ if the patient was still surviving after 30 days following surgery. This classification guarantees the outcome for each patient will be obtained in 31 days, which is much more practical.

We used the data of all 6994 patients to fit the following proportional odds logistic regression model,

$$\log\left(\frac{\pi_0(s)}{\pi_1(s)+\pi_2(s)}\right) = 3.637 - 0.074s, \ \log\left(\frac{\pi_0(s)+\pi_1(s)}{\pi_2(s)}\right) = 3.857 - 0.074s. \tag{6}$$

So the in-control probability of obtaining the outcome k, $\pi_k(s)$, $k = 0, 1, 2$, for a patient with risk score s using the fitted logistic regression model can be estimated as

$$\begin{cases} \pi_0(s) = \exp(3.637 - 0.074s)/\left[1+\exp(3.637 - 0.074s)\right] \\ \pi_2(s) = 1/\left[1+\exp(3.857 - 0.074s)\right] \\ \pi_1(s) = 1 - \pi_0(s) - \pi_2(s) \end{cases}. \tag{7}$$

For simplicity, we only examine the one-sided chart for detecting process deterioration. Thus, we set the in-control odds ratios to be $R_0 = R_1 = 1$ and the out-of-control values to be $R_0 = R_1 = 0.5$, respectively. Then the probability of obtaining the outcome k, $\pi_k^*(s)$, $k = 0, 1, 2$, for a patient with risk score s under the out-of-control case can be estimated by Equation (5). The CUSUM weight and statistic for patient t can be calculated using Equations (4) and (1), respectively. We set the constant control limit h to be 3.5 for illustration.

Tian *et al.* (2015) examined the in-control performance of the risk-adjusted Bernoulli CUSUM charts for five patient populations which were all patients ("All"), patients with higher 50% of Parsonnet scores ("High Risk"), patients with lower 50% of Parsonnet scores ("Low

Risk"), patients of surgeon 1 ("Surgeon 1") and patients of surgeon 6 ("Surgeon 6"). We used the same criteria to separate the 6994 patients into five populations in our simulation study. Since the Parsonnet score is the only explanatory variable in the proportional odds logistic regression model to determine the probability of different outcomes, we can use the different Parsonnet score distributions to represent the different patient populations. The histograms corresponding to the five Parsonnet scores distributions are displayed in Figure 5.1. It can be seen that the distributions of risk scores vary considerably. The mean Parsonnet scores for these five patient populations are listed in Table 5.1. If the varying patient populations have no significant impact on the chart performance, comparable in-control ARLs should be obtained for different risk distributions for the same control limits. In our study, 10,000 run lengths were obtained by randomly sampling from each of the five different risk distributions with replacement, and the in-control ARLs were estimated.

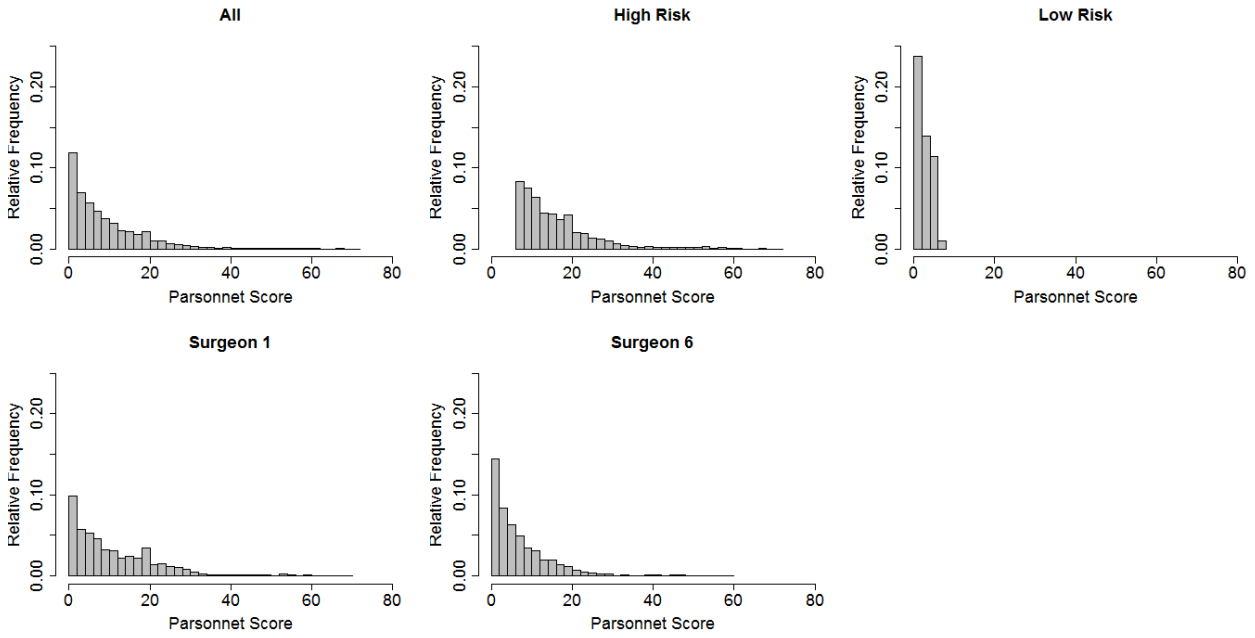


Figure 5.1. Histograms of Parsonnet scores for the five patient populations

The simulation results are listed in Table 5.1. The values in the parentheses are the standard errors of the estimated ARLs. It can be seen that the in-control ARLs vary significantly for different patient populations. More specifically, the in-control ARLs can vary by more than a factor of two under these realistic scenarios. In addition, there is an obvious decreasing trend in the in-control ARLs as the average Parsonnet score increases, as illustrated in Figure 5.2. The decreasing trend in the in-control ARLs as the mean of Parsonnet scores increases indicates an increase in false alarm rates when the average Parsonnet score gets higher. The results are similar to those of the study of Tian *et al.* (2015) on the risk-adjusted Bernoulli CUSUM chart. In summary, there is a significant impact of varying patient populations on the in-control performance of risk-adjusted CUSUM charts for multiresponses employing constant control limits. Therefore, we show in Section 3 that the control limits should be varying over time conditioning on the particular patient sequence in order to yield the desired consistent in-control performance regardless of the patient population. This makes the patient population irrelevant with respect to the CUSUM chart performance.

Table 5.1. Mean Parsonnet scores and estimated in-control ARLs comparison

Patient Populations	Mean Parsonnet score	Upper $\widehat{\text{ARL}_0}$(S.E.) with $h^+ = 3.5$
All	9.5193	2421.3 (23.42)
High Risk	16.4813	1694.0 (16.31)
Low Risk	2.5573	4085.1 (38.66)
Surgeon 1	11.3761	2156.0 (21.03)
Surgeon 6	7.4315	2768.8 (27.07)

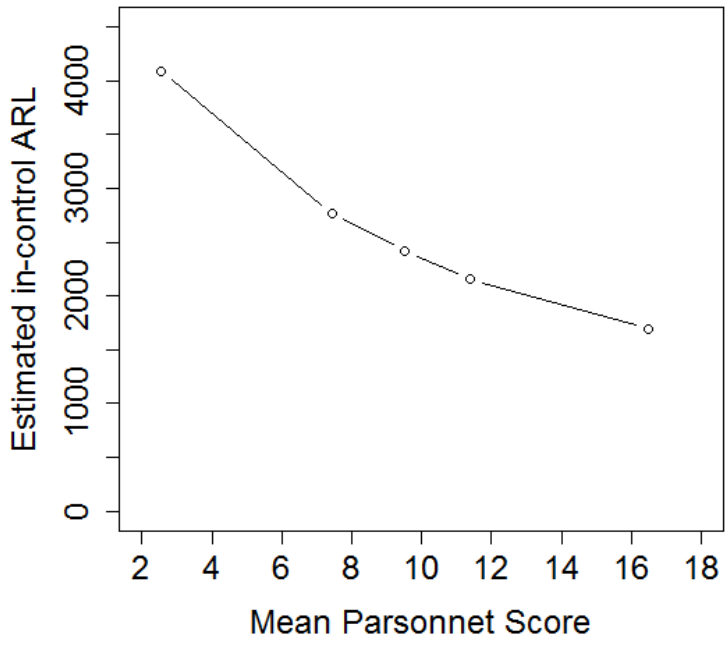


Figure 5.2. The estimated in-control ARLs of upper risk-adjusted CUSUM charts for multiresponses with constant control limits given varying patient populations

3. Dynamic probability control limits (DPCLs)

Zhang and Woodall (2015) recently developed a simulation-based method to determine the dynamic probability control limits (DPCLs) for risk-adjusted Bernoulli CUSUM charts based on the method of Shen *et al.* (2013). The Shen *et al.* (2013) method was developed for an application involving the exponentially weighted moving average (EWMA) chart to monitor Poisson count data with time-varying population sizes. The main idea is to keep the probability of obtaining a false alarm constant from sample to sample conditional on no false alarm for the previous observations, a concept initially proposed by Margavio *et al.* (1995). In our study, we apply the DPCLs to the risk-adjusted CUSUM charts based on multiresponses introduced in Section 2.1. The charts are henceforth referred to as risk-adjusted CUSUM charts for multiresponses with DPCLs. Specifically, for the upper risk-adjusted CUSUM chart based on multiresponses, the DPCLs $\boldsymbol{h}(\alpha)=\left(h_1(\alpha),\ h_2(\alpha),\ ...,\ h_k(\alpha),\ ...\right)$ satisfy to the extent possible the following equations:

$$\begin{cases} \Pr\left(C_1^+ > h_1(\alpha) \middle| S_1\right) = \alpha, \\ \Pr\left(C_t^+ > h_t(\alpha) \middle| C_k^+ \le h_k(\alpha),\ k=1,\ ...,\ t-1,\ S_t\right) = \alpha, \text{ for } t=2,\ 3,\ \ldots, \end{cases} \tag{8}$$

where $S_1,\ S_2,\ \ldots$ is the sequence of observed risk scores, such as Parsonnet scores, and α is the predetermined conditional false alarm rate. For each patient, the observed risk-adjusted CUSUM statistic $C_t^+ = \max\left(0, C_{t-1}^+ + W_t\right)$ is calculated, where W_t is computed using Equation (4) based on the observed $Y_t \in \{0,\ 1,\ \ldots,\ J\}$ and the risk score $S_t = s_t$ of the patient. The monitoring process signals when $C_t^+ > h_t(\alpha)$. From Equation (8), it follows that the in-control run length has

approximately a geometric distribution with parameter $p = \alpha$. Thus, the in-control ARL with use of the DPCLs $\boldsymbol{h}(\alpha) = (h_1(\alpha),\ h_2(\alpha),\ ...,\ h_k(\alpha),\ ...)$ should be close to $1/\alpha$.

The algorithmic form of the simulation procedure to obtain the control limit for patient t (t = 1, 2, 3, …) with risk score $S_t = s_t$, can be summarized as follows:

1) Generate N independent categorically distributed random variables $Y_{t,i}$ (i = 1, 2, …, N) $\in \{0,\ 1,\ \ldots,\ J\}$ with estimated in-control failure rates $(\pi_0(s_t),\ \ldots,\ \pi_J(s_t))$ = $(\pi_0^0(s_t),\ \ldots,\ \pi_J^0(s_t))$ obtained from risk-adjustment model (3). Calculate the N CUSUM statistics $C_{t,i} = \max(0, C_{t-1,j} + W_{t,i})$ (j = 1, 2, …, N) using the Equations (1) and (4) accordingly, where $C_{t-1,j}$ is randomly selected from a vector of CUSUM values $\boldsymbol{C}_{t\text{-}1}{}'$ $= \{C_{t-1,i}$ such that $C_{t-1,i} \le h_{t-1}(\alpha)\}$.

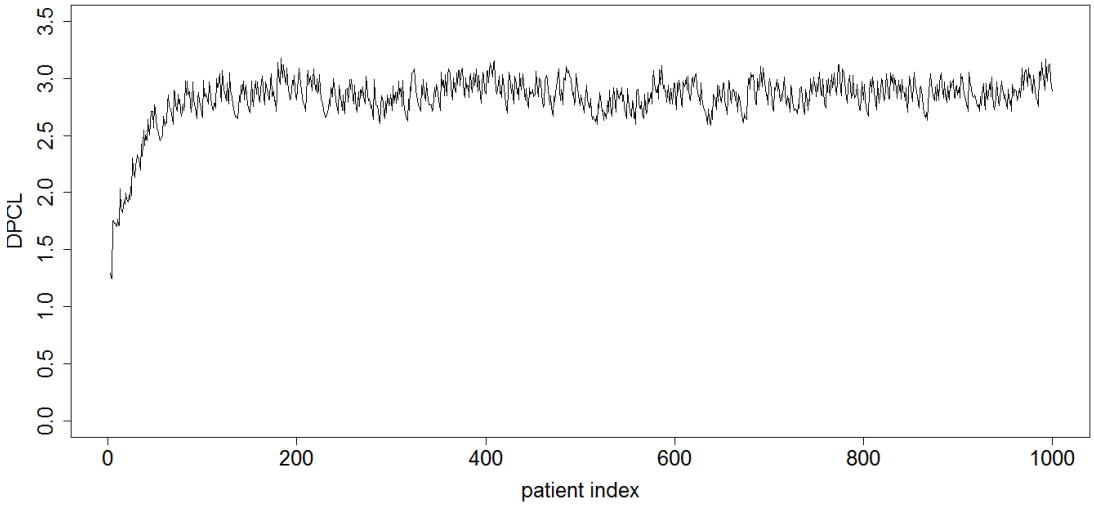
2) Sort the N CUSUM statistics in ascending order $C_{t,(1)}$, $C_{t,(2)}$, $C_{t,(3)}$, …, $C_{t,(N)}$ and take the $N' = \lceil N \cdot (1-\alpha) \rceil$ largest CUSUM statistic $C_{t,(N')}$ as the approximated DPCL $h_t(\alpha)$ if $\Pr\{C_{t,i} > C_{t,(N')}\} \neq 0$.

3) Calculate $C_t^+ = \max(0, C_{t-1}^+ + W_t)$ based on the observed surgical outcome and risk score of patient t. If $C_t^+ > h_t(\alpha)$, an out-of-control signal is issued. Otherwise, go back to step 1 and consider the next patient.

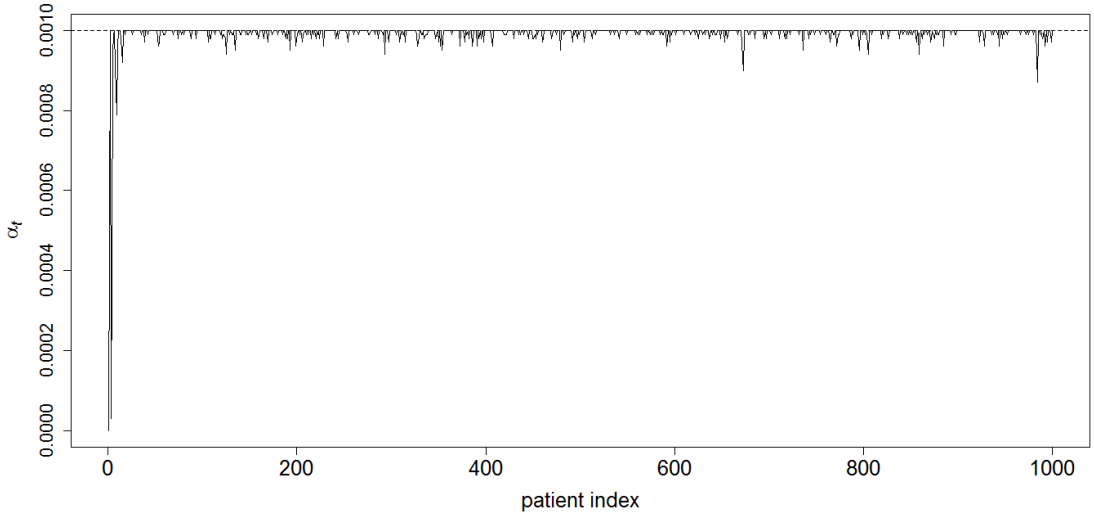
Due to the discreteness of categorically distributed random variables, for each patient t = 1, 2, …, $\Pr\{C_{t,i} > h_t(\alpha)\}$ (denoted by α_t) is always controlled to be less than or equal to α and $\Pr\{C_{t,i} \ge h_t(\alpha)\}$ (denoted by α_t') to be always greater than or equal to α. As the control limit

determination procedure progresses, an increasing number of different CUSUM statistic values will be generated, so the effect of the discreteness is somewhat higher at the early part of monitoring. For the first few patients there may not be a control limit such that the approximated conditional false alarm rate $\alpha_t = \Pr\{C_{t,i} > h_t(\alpha)\}$ is less than or equal to α. In these cases, there is no dynamic probability control limit used and no possibility of the chart signaling. After a relatively small number of patients, we will observe more $\alpha_t = \Pr\{C_{t,i} > h_t(\alpha)\}$ values close to α, but variability remains.

Figure 5.3 gives an example of DPCLs and the corresponding conditional false alarm rates $\alpha_t = \Pr\{C_{t,i} > h_t(\alpha)\}$ for the first 1000 patients in a sequence randomly chosen from the patient population "Surgeon 1". Here the conditional false alarm rate α was set as 0.001. Figure 5.3(a) shows the DPCLs obtained by applying the proposed procedure. It can be seen that the probability control limits vary dynamically from patient to patient. Figure 5.3(b) illustrates that the conditional false alarm rate $\alpha_t = \Pr\{C_{t,i} > h_t(\alpha)\}$ for each patient *t* is always controlled to be less than or equal to α = 0.001. Thus, one can expect to obtain approximately geometrically distributed in-control run lengths by applying DPCLs to the risk-adjusted CUSUM charts based on multiresponses.



(a) DPCLs



(b) $\alpha_t = \Pr\left\{C_{t,i} > h_t(0.001)\right\}$

Figure 5.3. (a) DPCLs, and (c) α_t = Pr{$C_{t,i}$ > h_t(0.001)} of the first 1000 patients in a sequence from risk distribution "Surgeon 1"

4. Simulation study

4.1. Simulation settings

In our simulation study, we assessed the in-control performance of the risk-adjusted CUSUM chart for multiresponses with DPCLs for specific sequences of patients from different patient populations. We used the same data and five patient populations (*i.e.* risk distributions) examined in Section 2.2. The same in-control risk-adjustment model (6) was used. We randomly chose two sequences of 20,000 Parsonnet scores with replacement from each of the five different risk distributions. Then for each sequence of scores, we calculated the DPCLs by applying the algorithm explained in Section 3. The conditional false alarm rate α was chosen to be 0.001 and N, the number of CUSUM statistics simulated, was set to be 100,000. The risk-adjusted CUSUM statistics were obtained using Equations (1), (4) and (5) with R_A = 0.5. We also recorded $\alpha_t = \Pr\{C_{t,i} > h_t(\alpha)\}$ and $\alpha_t' = \Pr\{C_{t,i} \geq h_t(\alpha)\}$ for t = 1, 2, … as a check for accuracy of the procedure. After we obtained a sequence of DPCLs for a specific sequence of scores selected from a particular risk distribution, 10,000 control charts were simulated to estimate the in-control ARL (ARL_0). Also, the 10th, 25th, 50th, 75th, 90th percentiles of the run length ($Q_{0.10}$, $Q_{0.25}$, $Q_{0.50}$, $Q_{0.75}$, $Q_{0.90}$) and the standard deviation of the in-control run length ($SDRL_0$) were estimated in order to compare with those of the geometric distribution with p = 0.001.

4.2. Estimated in-control performance

The simulation results are listed in Table 5.2 along with the parameters of the geometric distribution with p = 0.001 for comparison. From Equation (8), theoretically, the in-control run length of the risk-adjusted CUSUM chart based on multiresponses with DPCLs for any sequence

of patients should be approximately geometrically distributed with the parameter p equal to the specified conditional false alarm rate α. It can be seen that the estimated ARL_0, $SDRL_0$ and $Q_{0.10}$, $Q_{0.25}$, $Q_{0.50}$, $Q_{0.75}$, $Q_{0.90}$ values for each sequence examined are close to the corresponding theoretical values of a geometric distribution with p = 0.001. As it would be for the geometric distribution, the estimated $SDRL_0$ value is approximately equal to the estimated ARL_0 value for each sequence. However, we do observe slight deviations of the run length distributions from the geometric distribution, which is due to the discrete nature of the categorically distributed random variables. The average value of $\Pr\{C_{t,i} > h_t(0.001)\}$ (denoted by $\overline{\alpha_t}$) and the average value of $\Pr\{C_{t,i} \geq h_t(0.001)\}$ (denoted by $\overline{\alpha_t'}$) are also reported for each sequence of scores. We can see that the $\overline{\alpha_t}$ and $\overline{\alpha_t'}$ values are close to the desired conditional false alarm rate of α = 0.001. The same consistency as it pertains to the in-control performance is expected for other values of desired conditional false alarm rate α when DPCLs are applied to the risk-adjusted CUSUM charts based on multiresponses. The simulation results show that the DPCL method has consistent performance for any patient mix. By using DPCLs, one can obtain a desired in-control run length distribution close to the geometric distribution for any patient sequence from any risk distribution.

We also compared the DPCLs of two sequences from two risk distributions ("High Risk" and "Low Risk") with the largest difference in mean Parsonnet scores in Figure 5.4. Recall from Table 5.1 that the in-control performance is quite different for these two risk distributions when using a constant control limit. Figure 5.4 shows that different risk distributions can lead to quite different sets of dynamic probability control limits, but both sets of DPCLs resulted in desirable in-control performance. Using the DPCLs approach we were able to obtain similar in-control

ARLs (994.8 and 986.3) and average conditional false alarm rates of 9.9673E-4 and 9.9755E-4, respectively.

Table 5.2. Estimated in-control performance of risk-adjusted CUSUM charts for multiresponses with DPCLs (α = 0.001)

Risk distribution	Sequence index	$\widehat{\mathrm{ARL}_0}$(S.E.)	$\widehat{\mathrm{SDRL}_0}$	$Q_{0.10}$	$Q_{0.25}$	$Q_{0.50}$	$Q_{0.75}$	$Q_{0.90}$	$\overline{\alpha_t}$	$\overline{\alpha_t'}$
All	1	991.2 (9.85)	985.1	108	293	689	1375	2250	9.9510E-4	1.0173E-3
	2	1005.9 (9.94)	994.0	114	288	698	1400	2287	9.9498E-4	1.0161E-3
High Risk	1	994.8 (10.11)	1011.6	103	279	677	1376	2282	9.9673E-4	1.0187E-3
	2	1001.0 (9.98)	998.2	108	288	696	1416	2277	9.9668E-4	1.0154E-3
Low Risk	1	986.3 (9.79)	979.2	110	290	683	1366	2265	9.9755E-4	1.0139E-3
	2	994.1 (9.91)	991.1	108	281	689	1372	2301	9.9752E-4	1.0147E-3
Surgeon 1	1	1002.3 (9.86)	986.4	111	296	699	1408	2276	9.9521E-4	1.0166E-3
	2	1015.2 (10.28)	1027.7	109	288	711	1394	2320	9.9530E-4	1.0171E-3
Surgeon 6	1	987.3 (9.93)	993.4	109	285	682	1353	2272	9.9544E-4	1.0169E-3
	2	1009.9(9.98)	997.9	118	301	713	1397	2290	9.9528E-4	1.0171E-3
Geometric	/	1000 (/)	1000	105	287	692	1385	2301	/	/

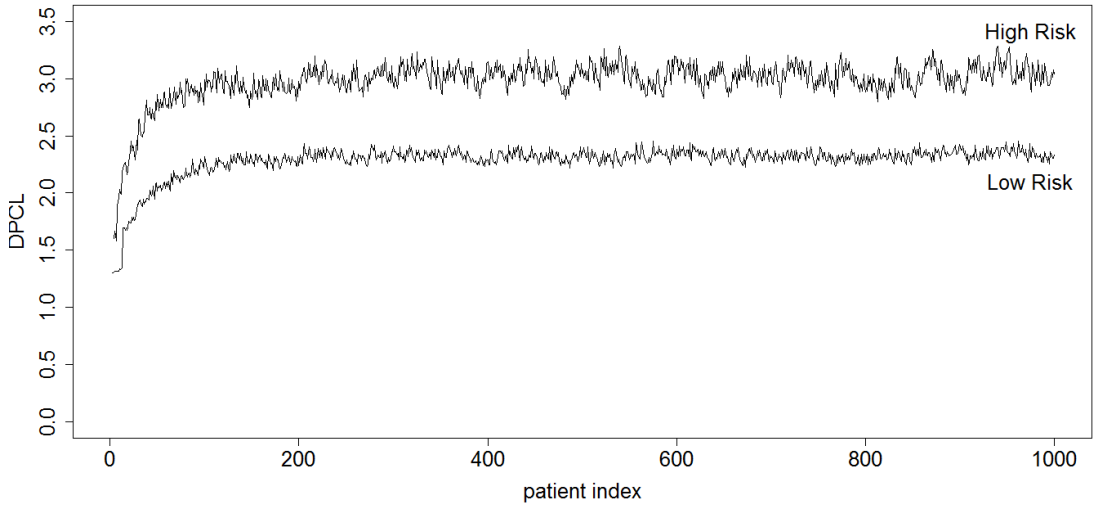


Figure 5.4. DPCLs comparison of two sequences from two different risk distributions ("High Risk" and "Low Risk")

If the risk distribution changes over time, as it would for a trainee surgeon that gains more experience, then the DPCLs approach will result in appropriate limit adjustments to accurately reflect the change. This is illustrated in Figure 5.5 where the population shifts from the lower risk patients to the higher risk patients after patient 500. The DPCLs adjust automatically after the 500th patient to higher values to accommodate the change of risk distribution. It is not necessary to estimate the new risk distribution as recommended by Tang *et al.* (2015).

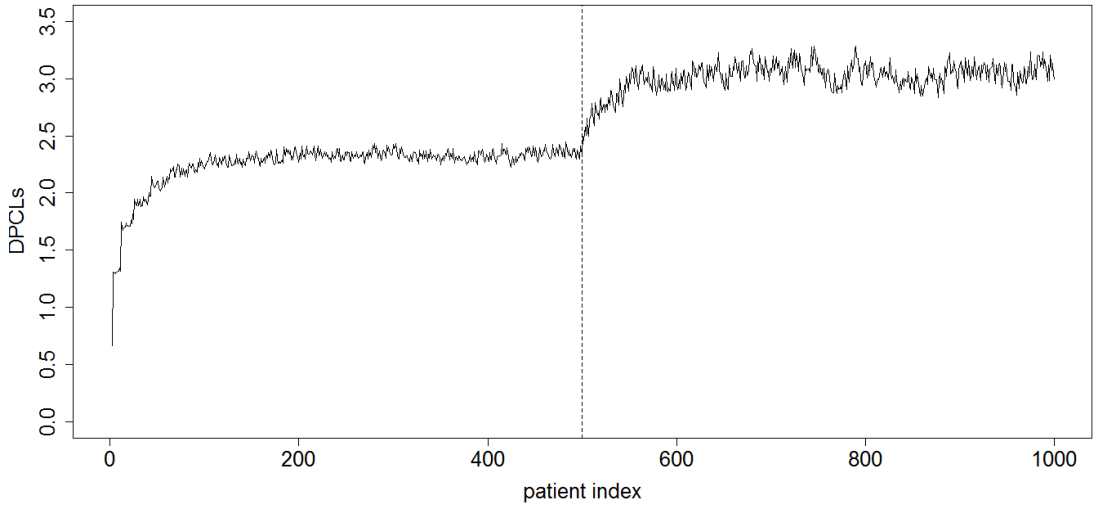


Figure 5.5. DPCLs where the risk distribution shifted from "Low Risk" to "High Risk" after the 500th patient

4.3. Comparison of constant control limit and DPCLs

From Table 5.1, the in-control ARL of the upper risk-adjusted CUSUM chart based on multiresponses for the risk distribution "All" is 2421.3 when using the constant control limit h^+ = 3.5 under the assumption that the patient population is known. To compare the constant control limit with DPCLs, we set the false alarm rate at 1/2421.3 = 0.000413 to obtain the similar in-

control ARL. The results are shown in Figure 5.6. The dashed line represents the constant control limit $h^+ = 3.5$, while the solid line indicates the DPCLs for a particular sequence of scores randomly chosen from the risk distribution "All". The two types of control limits result in very similar in-control ARL performance for the risk-adjusted CUSUM chart based on multiresponses. The advantage of DPCLs is that one can obtain the same in-control ARL and maintain the conditional false alarm rates at the specified level without requiring any information or assumptions about the patient population. Also, it is possible to signal sooner for process changes that occur near the start of the monitoring process by applying DPCLs. With the DPCLs one could signal at the third patient whereas with the constant control limit a signal could not be given until the sixth patient. Figure 5.6 would lead one to believe that steady-state out-of-control performance would be quite similar if the risk distribution assumed for the use of constant control limits were correct.

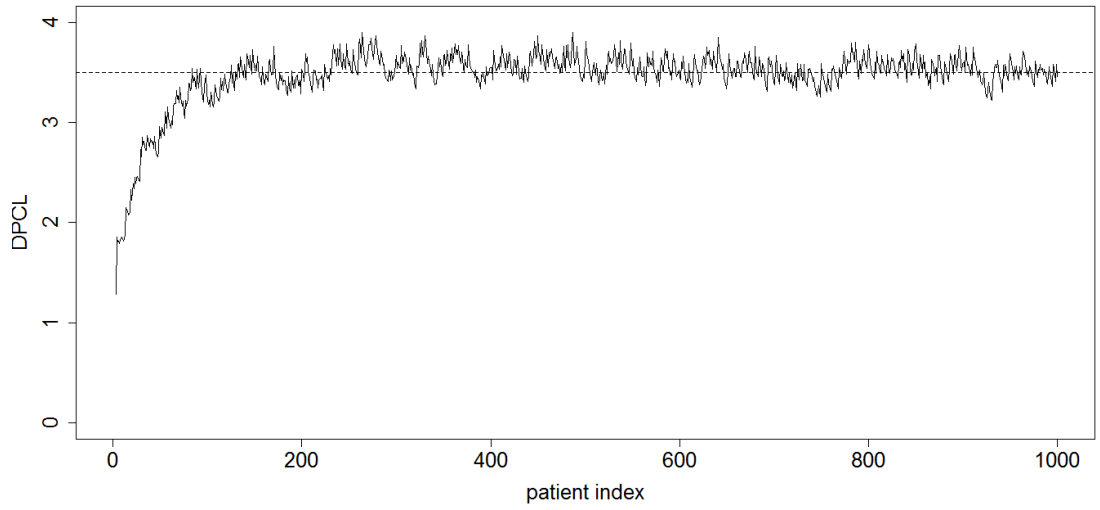


Figure 5.6. Comparison of constant control limit (dashed line) and DPCLs (solid line) for comparable in-control ARLs (α = 0.000413)

In addition, we compared in Figure 5.7 the conditional false alarm rates for the two charts shown in Figure 5.6. It is clearly seen that the conditional false alarm rates of the risk-adjusted CUSUM chart with DPCLs (darker line) were controlled at a specified level α = 1/2421.3 = 0.000413 while the ones of the chart with the constant control limit (lighter line) varied considerably, an undesirable feature.

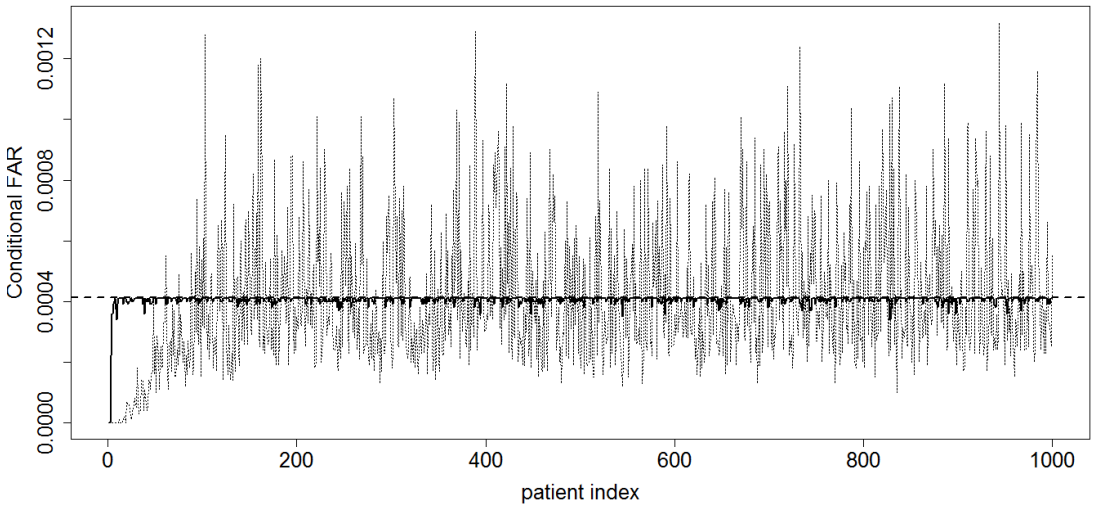


Figure 5.7. Comparison of the conditional false alarm rates (FARs) for the DPCLs control chart (darker line) and the ones for the constant limit control chart (lighter line) shown in Figure 5.6 (α = 0.000413)

5. Discussion

The risk-adjusted Bernoulli CUSUM chart developed by Steiner *et al.* (2000) is a popular tool for monitoring binary surgical outcomes. However, a patient surviving a surgery could have several levels of recovery and thus it is reasonable to consider more than two outcomes when monitoring the surgical outcome quality. Tang *et al.* (2015) developed the risk-adjusted CUSUM chart based on multiresponses for monitoring a surgical process with three or more mutually exclusive outcomes. Similar to the risk-adjusted Bernoulli CUSUM chart, there is a significant effect of varying risk distributions on the in-control performance of risk-adjusted CUSUM chart for multiresponses when constant control limits are applied. We found that the in-control ARLs of risk-adjusted CUSUM charts for multiresponses with constant control limits and a given risk-adjustment model can vary by a factor of two for different realistic patient populations. Thus, setting constant control limits for risk-adjusted CUSUM charts in different applications is not appropriate in practice. To address this problem, we applied the dynamic probability control limits proposed by Zhang and Woodall (2015) to the risk-adjusted CUSUM charts for multiresponses. By maintaining the conditional false alarm rate at a constant level given there is no false alarm for previous observations, the charts with DPCLs are able to give desirably consistent in-control performance with approximately geometrically distributed run lengths. Thus, the risk-adjusted CUSUM chart for multiresponses with DPCLs is more practical and should be applied to effectively monitor surgical performance by hospitals and healthcare practitioners.

The out-of-control performance of risk-adjusted CUSUM charts for multiresponses with DPCLs can be a future direction of study. The main purpose of applying DPCLs, however, is to

obtain desirably consistent in-control performance and control the conditional false alarm rates rather than to improve the detection ability of the chart. The method is computationally intensive and requires appropriate software program to be implemented. A less computationally intensive algorithm to determine DPCLs should be studied. Moreover, Zhang and Woodall (2015) applied DPCLs to the lower and two-sided risk-adjusted Bernoulli CUSUM charts and were able to obtain the desired in-control performance. It is reasonable to believe that the DPCL algorithm can be applied similarly to the lower side of a two-sided risk-adjusted CUSUM chart for multiresponses to detect improvement of surgical outcome quality. Finally, the effect of estimation error on the in-control performance of the charts with DPCLs is a topic to be investigated in future studies. Zhang and Woodall (2015) found that the effect of estimation error is uniformly smaller for the risk-adjusted Bernoulli CUSUM chart with DPCLs than for the corresponding chart with a constant control limit studied by Jones and Steiner (2012) under various realistic scenarios because the patients' risk distribution does not have to be estimated. One could expect the similar results for the charts based on multiresponses.

Acknowledgement

The work of W. H. Woodall was supported by National Science Foundation Grant CMMI-1436365.

References

Beiles, C. B., and Morton, A. P. (2004), "Cumulative Sum Control Charts for Assessing Performance in Arterial Surgery," *ANZ Journal of Surgery*, 74, 146-151.

Bottle, A., and Aylin, P. (2008), "Intelligent Information: A National System for Monitoring Clinical Performance," *Health Services Research*, 43, 10-31.

Chen, T. T., Chung, K. P., Hu, F. C., Fan, C. M., and Yang, M. C. (2011), "The Use of Statistical Process Control (Risk-Adjusted CUSUM, Risk-Adjusted RSPRT and CRAM with Prediction Limits) For Monitoring the Outcomes of Out-of-Hospital Cardiac Arrest Patients Rescued by the EMS System," *Journal of Evaluation in Clinical Practice*, 17, 71-77.

Collins, G. S., Jibawi, A., and McCulloch, P. (2011), "Control Chart Methods for Monitoring Surgical Performance: A Case Study from Gastro-Oesophageal Surgery," *European Journal of Surgical Oncology (EJSO)*, 37, 473-480.

Cook, D. A., Duke, G., Hart, G. K., Pilcher, D., and Mullany, D. (2008), "Review of the Application of Risk-Adjusted Charts to Analyze Mortality Outcomes in Critical Care," *Critical Care and Resuscitation: Journal of the Australasian Academy of Critical Care Medicine*, 10, 239-251.

Grigg, O., and Farewell, V. (2004), "An Overview of Risk-Adjusted Charts," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 167, 523-539.

Harris, J. R., Forbes, T. L., Steiner, S. H., Lawlor, D. K., DeRose, G., and Harris, K. A. (2005), "Risk-Adjusted Analysis of Early Mortality after Ruptured Abdominal Aortic Aneurysm Repair," *Journal of Vascular Surgery*, 42, 387-e1.

Hawkins, D. M., Qiu, P., and Kang, C. W. (2003), "The Changepoint Model for Statistical Process Control," *Journal of Quality Technology*, 35, 355-366.

Loke, C. K., and Gan, F. F. (2012), "Joint Monitoring Scheme for Clinical Failures and Predisposed Risks," *Quality Technology and Quantitative Management*, 9, 3-21.

Lovegorove, J., Valencia, O., Treasure, T., Sherlaw-Johnson, C., and Gallivan, S. (1997), "Monitoring the Results of Cardiac Surgery by Variable Life-Adjusted Display," *Lancet*, 350, 1128-1130.

Jones, M. A., and Steiner, S. H. (2012), "Assessing the Effect of Estimation Error on Risk-Adjusted CUSUM Chart Performance," *International Journal for Quality in Health Care*, 24, 176-181.

Margavio, T. M., Conerly, M. D., Woodall, W. H., and Drake, L. G. (1995), "Alarm Rates for Quality Control Charts," *Statistics & Probability Letters*, 24, 219-224.

McCullagh, P. (1980), "Regression Models for Ordinal Data," *Journal of the Royal Statistical Society*, Series B, 42, 109-142.

Moore, R., Nutley, M., Cina, C. S., Motamedi, M., Faris, P., and Abuznadah, W. (2007), "Improved Survival after Introduction of an Emergency Endovascular Therapy Protocol for Ruptured Abdominal Aortic Aneurysms," *Journal of Vascular Surgery*, 45, 443-450.

Morton, A. P., Clements, A. C., Doidge, S. R., Stackelroth, J., Curtis, M., and Whitby, M. (2008), "Surveillance of Healthcare-Acquired Infections in Queensland, Australia: Data and Lessons from the First 5 Years," *Infection Control and Hospital Epidemiology*, 29, 695-701.

Novick, R. J., Fox, S. A., Stitt, L. W., Forbes, T. L., and Steiner, S. H. (2006), "Direct Comparison of Risk-Adjusted and Non–Risk-Adjusted CUSUM Analyses of Coronary Artery Bypass Surgery Outcomes," *The Journal of Thoracic and Cardiovascular Surgery*, 132, 386-391.

Page, E. S. (1954), "Continuous Inspection Schemes," *Biometrika*, 41, 100-115.

Parsonnet, V., Dean, D., and Bernstein, A. D. (1989), "A Method of Uniform Stratification of Risk for Evaluating the Results of Surgery in Acquired Adult Heart Disease," *Circulation*, 79, I3-I12.

Rogers, C. A., Reeves, B. C., Caputo, M., Ganesh, J. S., Bonser, R. S., and Angelini, G. D. (2004), "Control Chart Methods for Monitoring Cardiac Surgical Performance and Their Interpretation," *The Journal of Thoracic and Cardiovascular Surgery*, 128, 811-819.

Ryan, A. G., Wells, L. J., and Woodall, W. H. (2011), "Methods for Monitoring Multiple Proportions When Inspecting Continuously," *Journal of Quality Technology*, 43, 237-248.

Shen, X., Zou, C., Jiang, W., and Tsung, F. (2013), "Monitoring Poisson Count Data with Probability Control Limits When Sample Sizes Are Time Varying," *Naval Research Logistics*, 60, 625-636.

Sherlaw-Johnson, C. (2005), "A Method for Detecting Runs of Good and Bad Clinical Outcomes on Variable Life-Adjusted Display (VLAD) Charts," *Health Care Management Science*, 8, 61-65.

Sherlaw-Johnson, C., Morton, A., Robinson, M. B., and Hall, A. (2005), "Real-Time Monitoring of Coronary Care Mortality: A Comparison and Combination of Two Monitoring Tools," *International Journal of Cardiology*, 100, 301-307.

Sherlaw-Johnson, C., Wilson, P., and Gallivan, S. (2007), "The Development and Use of Tools for Monitoring the Occurrence of Surgical Wound Infections," *Journal of the Operational Research Society*, 58, 228-234.

Shortell, S. M., Jones, R. H., Rademaker, A. W., Gillies, R. R., Dranove, D. S., Hughes, E. F. X., Budetti, P. P., Reynolds, K. S. E. and Huang, C. F. (2000), "Assessing the Impact of Total Quality Management and Organizational Culture on Multiple Outcomes of Care for Coronary Artery Bypass Graft Surgery Patients," *Medical Care*, 38, 207-217.

Steiner, S. H. (2014), "Risk-Adjusted Monitoring of Outcomes in Health Care," Chapter 14 in *Statistics in Action: A Canadian Outlook*, Lawless, J.F., ed., pp. 245–264. Chapman and Hall/CRC.

Steiner, S. H., Cook, R. J., and Farewell, V. T. (2001), "Risk-Adjusted Monitoring of Binary Surgical Outcomes," *Medical Decision Making*, 21, 163-169.

Steiner, S. H., Cook, R. J., Farewell, V. T., and Treasure, T. (2000), "Monitoring Surgical Performance Using Risk-Adjusted Cumulative Sum Charts," *Biostatistics*, 1, 441-452.

Tang, X., Gan, F. F., and Zhang, L. (2015), "Risk-Adjusted Cumulative Sum Charting Procedure Based on Multiresponses," *Journal of the American Statistical Association*, 110, 16-26.

Tian, W., Sun, H., Zhang, X., and Woodall, W. H. (2015), "The Impact of Varying Patient Populations on the In-Control Performance of the Risk-Adjusted CUSUM Chart," *International Journal for Quality in Health Care*, 27, 31-36.

Woodall, W. H. (2006), "The Use of Control Charts in Health-Care and Public-Health Surveillance," *Journal of Quality Technology*, 38, 89-104.

Woodall, W. H., Fogel, S. L., and Steiner, S. H. (2015), "The Monitoring and Improvement of Surgical Outcome Quality," *Journal of Quality Technology*, 47, 383-398.

Zhang, X. and Woodall, W. H. (2015), "Dynamic Probability Control Limits for Risk-Adjusted Bernoulli CUSUM Charts," *Statistics in Medicine*, 34, 3336-3348.

Zhang, X. and Woodall, W. H. (2015), "Dynamic Probability Control Limits for Lower and Two-Sided Risk-Adjusted Bernoulli CUSUM Charts," Submitted for publication.

Zhang, X. and Woodall, W. H. (2015), "The Effect of Estimation Error on In-Control Performance of Risk-Adjusted Bernoulli CUSUM Chart with Dynamic Probability Control Limits," Submitted for publication.

Chapter 6

Conclusions and future work

The risk-adjusted Bernoulli CUSUM chart with constant control limits developed by Steiner *et al.* (2000) has become an increasingly popular tool for monitoring clinical and surgical outcome performance. However, its design requires information or assumptions about patient populations which are often unavailable or inaccurate. Previous studies showed that varying risk distributions of patients has a significant effect on the in-control performance of risk-adjusted Bernoulli CUSUM charts. Thus, setting constant control limits for risk-adjusted Bernoulli CUSUM charts in different applications is not suitable in practice. To overcome this disadvantage, we applied dynamic probability control limits (DPCLs) to the risk-adjusted Bernoulli CUSUM charts. By maintaining the probability of a false alarm at a constant level conditional on no false alarm for previous observations, the charts with DPCLs gave desirably consistent in-control performance with approximately geometrically distributed run lengths. The simulation results illustrated that our method does not rely on any information or assumptions about the patient populations or their risk distributions. In addition, one could run the upper and lower risk-adjusted Bernoulli CUSUM charts with DPCLs side by side and obtain the desired in-control performance for the two-sided chart. Thus, the risk-adjusted Bernoulli CUSUM chart with DPCLs is more practical and should be applied to appropriately monitor surgical performance by hospitals and healthcare practitioners. For future researches, less computationally intensive algorithms to determine DPCLs should be studied. For example, a Markov chain model was used to compute the DPCLs for the EWMA charts studied by Shen *et al.* (2013).

Next, we examined the effect of estimation error on the in-control performance of risk-adjusted Bernoulli CUSUM charts with DPCLs. The simulation results demonstrated that the in-control performance of risk-adjusted Bernoulli CUSUM chart with DPCLs is less affected by estimation error than the chart with a constant control limit. The most influential factors are the specified desired in-control average run length, the Phase I sample size and the overall adverse event rate. These finding are similar to the study of Jones and Steiner (2012). Although the $AARL_0$ can be controlled at the desired level, the variation among ARL_0's due to estimation error is not negligible. Therefore, updating the in-control risk-adjustment model was suggested to alleviate this problem. However, it was found that the effect of estimation error is uniformly smaller for the risk-adjusted Bernoulli CUSUM chart with DPCLs than the chart with a constant control limit under various realistic scenarios. The better performance is due to the fact that applying DPCLs does not require the estimation of the patient mix. Thus, use of DPCLs is demonstrated to have another advantage when designing a risk-adjusted Bernoulli CUSUM chart. Ways to reduce the effect of estimation error on the in-control performance of the risk-adjusted Bernoulli CUSUM chart with DPCLs should be the future direction of studies. For instance, using a bootstrap method to determine the estimated surgical failure rates and the control limits could be considered (Jones and Steiner (2012)).

Moreover, we adapted the DPCL approach to the risk-adjusted CUSUM chart based on multiresponses developed by Tang *et al.* (2015) for monitoring a surgical process with more than two outcomes since a patient surviving a surgery could have several levels of recovery. Similar to the risk-adjusted Bernoulli CUSUM chart, it was found that there is a significant effect of varying risk distributions on the in-control performance of risk-adjusted CUSUM chart for multiresponses when constant control limits are applied. The simulation results showed that the

charts with DPCLs are able to give desirably consistent in-control performance with approximately geometrically distributed run lengths. Therefore, the DPCL approach is proved to be adaptable for the risk-adjusted CUSUM chart for multiresponses. The out-of-control performance of risk-adjusted CUSUM charts for multiresponses with DPCLs can be a future direction of study. Also, the DPCL algorithm can be applied similarly to the lower side of a two-sided risk-adjusted CUSUM chart for multiresponses to detect improvement of surgical outcome quality. Finally, the effect of estimation error on the in-control performance of the risk-adjusted CUSUM charts for multiresponses with DPCLs is worth to be investigated in future studies.

References

(Not included in Chapters 2 – 5)

Zhang, X., Woodall, W. H., and Loda, J. (2015). Dynamic probability control limits for risk-adjusted CUSUM charts based on multiresponses. Submitted for publication.