Mathematical Biosciences and Engineering

MBE, 17(4): 3130–3146.
DOI: 10.3934/mbe.2020177
Received: 26 December 2019
Accepted: 24 March 2020
Published: 16 April 2020

http://www.aimspress.com/journal/MBE

Research article

A Fast Online Monitoring Approach for Surgical Risks

Jiaqi Liu [1], Xin Lai [1,*], Jiayin Wang [1], Paul B.S. Lai[2], Xuanping Zhang[1] and Xiaoyan Zhu[1]

[1] School of Computer Science and Technology, Xi'an Jiaotong University, Xi'an, China
[2] Department of Surgery, The Chinese University of Hong Kong, Hong Kong, China

* **Correspondence:** Email: laixin@xjtu.edu.cn.

Abstract: Risk monitoring has been widely used in health care, further, control charts are often used as monitoring methods for surgical outcomes. Most of the methods can only detect step shifts of position parameters, but cannot take measures on scale parameters. In this paper, we proposed four methods based on EWMA control charts, namely SESOP, STSSO, SESOP-MFIR and STSSO-MFIR, to improve the existing monitoring methods. Specifically, SESOP standardizes variable on the basis of an EWMA charting method; STSSO replaces the statistics of the original EWMA charting method with the score test statistics; for SESOP-MFIR and STSSO-MFIR, we upgrade their control limits from asymptotic to time-varying based on SESOP and STSSO, which enhance the timeliness of the earlier shifts monitoring. In order to verify the improvement of surgical outcomes monitoring, we respectively carry out simulation experiment and a practical application on ESOP and our four methods. SESOP can raise the overall efficiency of detecting shifts; STSSO led to a significant increase in the monitoring stability, especially for small volatilities; the optimization brought by SESOP-MFIR and STSSO-MFIR are more obvious, that the speed of detecting earlier shifts can even be reduced to half of the existing methods. Then, we apply these methods to the SOMIP program of Hong Kong, SESOP-MFIR and STSSO-MFIR have the best performance and can detect early shifts in time. According to the results, the methods we proposed can monitor both early shifts and scale parameters and improve the performance of surgical outcome monitoring in different degrees compared to those existing methods.

Keywords: EWMA charting method; risk monitoring; FIR features; variable standardization; score test statistics

Abbreviations

EWMA: Exponentially Weighted Moving Average; CUSUM: Cumulative Sum; SOMIP: Surgical Outcome Monitoring and Improvement and Improvement Program; VLAD: Variable Life-Adjusted Display; ARL: Average Run Length; WST: Weighted Score Test; GLMMs: Generalized Linear Mixed Models; IC: In Control; OC: Out of Control; FIR: Fast Initial Response; CL: Control Limit; ESOP: EWMA chart for Surgical Outcome; SESOP: Standardization EWMA chart for Surgical Outcome; STSSO: Score Test Statistics for Surgical Outcome; SESOP-MFIR: Standardization EWMA chart for Surgical Outcome with Modified Fast Initial Response; STSSO-MFIR: Score Test Statistics for Surgical Outcome with Modified Fast Initial Response.

1. Introduction

In today's manageable health care environment, measuring and comparing the quality of health care across institutions is becoming increasingly important. So, there has been a great deal of interest in improving the quality of health care, with particular emphasis on monitoring surgical outcomes [1]. In the past few years, control charts have been widely used in surgical quality, which are able to help identify the source of the problems and provide ideas on how to solve the problems [2]. Control charting methods are statistically designed charts that measure, record, and evaluate process quality characteristics to monitor whether the process is in a controlled state, which can either be a memory-type or memory-less control chart [3], the representative methods are exponentially weighted moving average (EWMA) charts [4–6], cumulative sum (CUSUM) charts [7–9], Variable Life-Adjusted Display (VLAD) charts [10,11] and Shewhart charts [12–14]. Although the concept of control charts was proposed earlier, the subsequent developments are basically based on the improvement and comparison of the above control charts [3,15–17].

However, unlike the relative stability of industrial processes, the target of surgical risk monitoring is the patient, whose the postoperative state is largely affected by the preoperative physical health and surgical factors, such as age, dyspnea status, magnitude of surgery, and so forth [18]. To accommodate such heterogeneity, the risk-adjusted charts [19–21] were used to monitoring the risks, in which a risk adjustment procedure was employed. Specifically, Cook et al. adopted a risk-adjusted EWMA chart to sequentially evaluate the results in intensive care units [19]; Steiner et al. introduced the logit adjusted model into CUSUM chart in order to reflect the variation of surgical risk among patients [20]; Jin Y et al. proposed a new risk-adjusted exponentially weighted moving average VLAD chart and provided a control limit that can be used with the VLAD [21]; Zhang et al. determined simulation-based dynamic probability control limits (DPCLs) patient-by-patient for the risk-adjusted Bernoulli CUSUM charts [22]. On the whole, in the existing methods, EWMA chart is used to monitor surgical outcomes [1,23]; CUSUM chart has been proposed to detect the occurrence of postoperative deterioration [24,25]; VLAD chart is widely used as a tool to produce clear and easy-to-understand results in monitoring of surgical outcomes [10,11]. Most of the monitoring methods mentioned here focus on the detection of step shifts in position parameters, however, online monitoring of scale parameters is a significant part of surgical monitoring outcomes of surgery, but it is not covered by these methods [18]. Therefore, it is particularly important to have appropriate surgical risk monitoring methods and assist decision-making by hospital administrators and health care providers.

As we all know, the changes in both average risk and volatility are important characteristics that needs to be detected in monitoring surgical outcomes [18] and has a critical impact on survival. However, the above methods monitor the surgical outcomes in average levels, their alarm control limits are also fixed values, which means the same weight for all the moments in the monitoring process, and make them less effective in monitoring volatility of surgical performance. Back to the practical applications, in the annual surgical monitoring, medical system administrators and hospital staffs expect to detect abnormalities of surgical outcomes at the early stage of monitoring process, which can implement the intervention earlier to correct the identified problems. However, the existing surgical monitoring methods are difficult to respond to early changes in a timely manner.

In 2008, a program was launched from the Hospital Authority of Hong Kong named Surgical Outcome Monitoring and Improvement and Improvement Program (SOMIP), which annually audits the surgical performance of all public hospitals in the area and has significant effects in monitoring 30-day mortality risk and identifying deterioration. Specifically, SOMIP takes advantage of physical and biochemical factors like age and albumin as preoperative factors to estimate the postoperative surgery risk, and uses them as benchmark to measure the surgical performance [18]. Figure 1 is a VLAD chart for 30-day mortality of a hospital in SOMIP project over a four-year period (from 1 Jul 2009 to 30 Jun 2013), in which the vertical axis indicates the difference between the actual number of deaths and the expected number of death. It could be observed that the quality of this hospital started to deteriorate at the beginning. Some methods were proposed to online monitor such changes in surgical risks. Based on the in-control data that are collected in phase I, those approaches can be used in phase II to consecutively judge whether the risks significantly deviate to the in-control state as the cases occur sequentially. However, previous surgical monitoring methods may give a delay alert because it usually takes a certain time to trigger the alarm. Therefore, a more efficient monitoring method for early-stage changes could help to identify the surgical quality problems earlier.

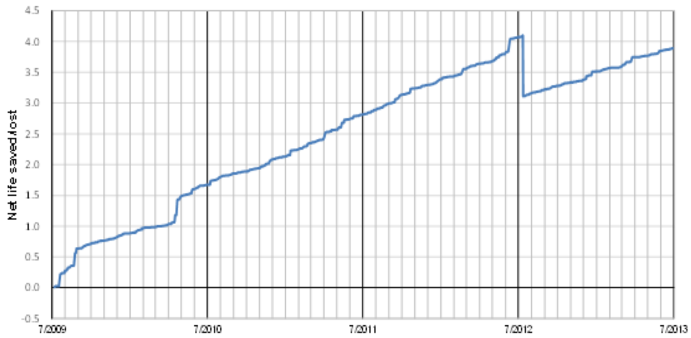

Figure 1. VLAD chart for 30 day mortality based on 4 year model.

Specifically, a shortcoming of both EWMA and CUSUM control charts is that they have asymptotic control limits, and using an asymptotic control limit rather than a time-varying limit may result in a slightly higher zero-state ARL that would make a monitoring chart less sensitive to start-up quality problems [26]. To enhance the sensitivity of the EWMA control chart to start-up quality problems, Steiner et al. [26] proposed a single EWMA control chart based on fast initial response (FIR) adjustment. In terms of the application maturity, Lucas and Crosier [27] proposed a fast initial response method for CUSUM control chart, which improved the sensitivity of monitoring volatilities of CUSUM control chart. In the following decades of application, the rationality of FIR method has been well developed [26,27,29]. In this article, on the one hand, we extend the work of Liu et al. (named ESOP) [18], propose a new monitoring chart based on ESOP and further improve the performance in monitoring surgical outcomes using a modified FIR adjustment; on the other hand, we propose a monitoring method based on a new weighted score test (WST) and further enhance the sensitivity to start-up quality problems.

The remainder of this article is organized as follows. Existing monitoring method ESOP and our proposed methods are introduced in Section 2. In Section 3, simulation studies and a real-world application are carried out to evaluate the performance of the proposed monitoring methods under different indicators. Finally, the results discussion and conclusion are shown in the remainder parts.

2. Materials and method

2.1. EWMA control statistic for surgical outcomes

Proposed by Liu et al. [18], a "risk-adjusted" method, which named ESOP, has been accounted for surgery's risk. ESOP is based on weighted score test (WST), which could be extended to a generalized condition where various types of outcomes modeled by generalized linear mixed models (GLMMs) can be monitored. Taking into account the actual situation of patient postoperative risk monitoring, a charting statistic Z_n is proposed,

$$Z_n = \sum_{i=1}^{t} \lambda(1-\lambda)^{n-i}[(y_i - p_i)^2 - p_i(1-p_i)] \tag{2.1}$$

where n is the current time point, i is the past time points, λ denotes smoothing factor, $\lambda(1-\lambda)^{n-i}$ is exponentially weighted, p_i denotes the surgical failure rate, y_i is independent binary surgical outcomes, when the surgical operation fails, $y_i = 1$, otherwise, $y_i = 0$. Equation (2.1) is equivalent to

$$Z_n = (1-\lambda)Z_{n-1} + \lambda Y_n, \quad n = 1,2,\ldots \tag{2.2}$$

with

$$Y_n = (y_n - p_n)^2 - p_n(1-p_n) \tag{2.3}$$

First, we construct a risk-adjusted model

$$logit(p_n) = x_n\beta + \alpha \tag{2.4}$$

where x_n is a sequence of independent and identically distributed normal random variables, which denotes the risk factor of nth patient y_n, α is an intercept parameter, β is another parameter, by equation (2.4) we can obtain $p_n = \frac{exp\,(x_n\beta+\alpha)}{1+exp(x_n\beta+\alpha)}$. In general, the above situation is called in control (IC), which means that y_n~Bernoulli(1, p_n). With the change of time, suppose y_n enter another state at a change-point ε, and the corresponding surgical failure rate becomes ${p_n}^*$, thus y_n~Bernoulli(1,${p_n}^*$). If all data follow an identical distribution ($p_n = {p_n}^*$), the process is IC; otherwise, the process enter out of control (OC) state, which makes more sense in monitoring surgical outcomes. Therefore, in order to achieve the purpose of evaluating the outcomes of surgery, we need to test the following hypotheses:

$$H_0: p_n{}^* = p_n \qquad \text{versus} \qquad H_1: p_n{}^* \neq p_n$$

Specifically, the reasons for $p_n{}^* \neq p_n$ are various [19], and two typical cases are selected for consideration in this paper. On the one hand, from $p_n = \frac{exp\,(x_n\beta+\alpha)}{1+exp(x_n\beta+\alpha)}$ (IC) to $p_n{}^* = \frac{exp\,(x_n\beta+\alpha+\delta)}{1+exp(x_n\beta+\alpha+\delta)}$ (OC) may be caused by shift of the location parameter; on the other hand, scale parameter's shift could cause a change in surgical failure rate, such as from $p_n = \frac{exp\,(x_n\beta+\alpha)}{1+exp(x_n\beta+\alpha)}$ (IC) to $p_n{}^* = \frac{exp\,(x_n\beta+\alpha+\omega)}{1+exp(x_n\beta+\alpha+\omega)}$ (OC), where ω is a random effect and can be defined as $\omega = \tau\vartheta$, τ denotes a constant, ϑ follows an unspecified distribution.

Then, we proposed an EWMA charting method (SESOP) to monitor outcomes of surgery, which standardizes Y_n on the basis of ESOP, and makes Y_n generally obey the standard normal distribution, thereby improving the performance of monitoring surgical outcomes, details as follows:

$$Y_n = \frac{(y_n-p_n)^2-p_n(1-p_n)}{\sqrt{p(1-p)(1-2p)^2}} \tag{2.5}$$

2.2. One-sided score-type test for surgical outcomes

Similarly to Silvapulle et al. [30], we use a test statistics T for $H_0: \sigma^2 = 0$ against $H_1: \sigma^2 > 0$ as

$$T = U_0{}^t\widetilde{D}^{-1}U_0 - inf\{(U_0 - b)^t\widetilde{D}^{-1}(U_0 - b): b > 0\} \tag{2.6}$$

According to the characteristics of surgical outcome monitoring, we made the following definitions:

$$U_0 = \frac{1}{2}[(y - U)^2 - V] \tag{2.7}$$

$$\widetilde{D} = \frac{1}{4}[(1 - 2U)^2V] \tag{2.8}$$

where $U = p$, $V = p(1 - p)$, p denotes the surgical failure rate. Similarly, the statistics Z_n can be expressed as

$$Z_n = (1 - \lambda)Z_{n-1} + \lambda T,\ n = 1,2, \dots \tag{2.9}$$

Note that using the values of time series to replace the variables in the above equations. Thereafter, we refer to the chart using the Z_n as the control statistics as STSSO to monitor surgical outcomes. For positive σ^2, the score in (2.6) at zero is positive, so that the infimum gets zero. When σ^2 is negative, the score at zero is also negative, so $T = 0$ [31].

2.3. FIR features for control charts

Through the above descriptions, we have proposed two EWMA control charts (SESOP and STSSO) to monitor surgical outcomes, which all perform well, especially when the data has small volatilities; however, the control limits of SESOP and STSSO are asymptotic control limits, which make them less sensitive to identify start-up problems. Back to the actual situation, the early volatilities are crucial in surgical risk monitoring. For example, the SOMIP in Hong Kong is officially evaluated once a year, it is also helpful to improve the quality of surgery if it can identify start-up problems as early as possible. To improve this problem, we introduce a fast initial response (FIR) feature, which is proposed by Lucas et al. [27] to enhance the performance of a control chart especially when the control chart gives an OC signal. Then, Steiner et al. [26] proposed a single EWMA control chart based on fast initial response (FIR) adjustment, which is named FIR_{adj} and given by

$$FIR_{adj} = 1 - (1 - f)^{1+a(t-1)} \tag{2.10}$$

and stated the EWMA control chart combined with FIR feature is more sensitive to early process volatilities, especially when the EWMA weight is small. On the basis of FIR adjustment (FIR_{adj}), Abdul et al. [28] further improved the performance of FIR_{adj} by using a power transformation with respect to time t, which is named modified FIR_{adj} ($MFIR_{adj}$) and given by

$$MFIR_{adj} = \left\{1 - (1 - f)^{1+a(t-1)}\right\}^{1+\frac{1}{t}} \tag{2.11}$$

Specifically, the function of the adjustment parameter a can reduce the influence of $MFIR_{adj}$ after time t; f is a proportion that determines the distance between the control limit of the first sample point and the starting value. As can be seen, MFIR feature uses a power transformation about time t compared to FIR feature, which helps in further decreasing the control limits and makes the EWMA control chart more sensitive when detecting earlier shifts.

We extend the work of Abdul et al. [28] and combine $MFIR_{adj}$ with SESOP and STSSO to turn asymptotic control limits into time-varying control limits. The SESOP and STSSO based on $MFIR_{adj}$ are named as SESOP-MFIR and STSSO-MFIR respectively. The control limits of SESOP-MFIR control chart is as follows:

$$CL_t = h\left\{1 - (1 - f)^{1+a(t-1)}\right\}^{1+\frac{1}{t}} \tag{2.12}$$

where h is the asymptotic control limit (CL) of SESOP. STSSO-MFIR is similar to SESOP-MFIR.

2.4. Detail description and method overview

Different from those methods that have the same weight for all data, the methods proposed in this paper based on EWMA chart assign different weights according to time series. In general, more recent observations receive more weight, which is also consistent with the temporal characteristic of the surgical outcomes data. In SESOP and STSSO, we define their control limits as the same asymptotic control limits as ESOP, in order to verify the improvement of the EWMA chart performance by standardization and one-sided score test. In SESOP-MFIR and STSSO-MFIR, we upgrade asymptotic control limits to time-varying control limits based on SESOP and STSSO, which improve the timeliness of the earlier shifts monitoring.

The status of the monitoring process includes IC and OC, the initial state is generally IC, which is called Phase I; after a period of time, it either keep in IC or change to OC, we call it Phase II. In Phase I, when we obtain an observation, we can get the chart statistics Z_n through $p_n =$

$\frac{exp\,(x_n\beta+\alpha)}{1+exp(x_n\beta+\alpha)}$ and equations (2.2)(2.3). In Phase II, when we obtain a new observation, we calculate the chart statistics Z_n similarly and compare it with the control limit. If Z_n exceeds the control limit, the monitoring system gives an OC signal, otherwise, it continues to monitor new observations. It can be seen that the difference between combining $MFIR_{adj}$ or not is whether the control limit in Phase II is a value changing with time or a fixed value. It is worth noting that the p_n's value corresponding to different n is not equal, it is intractable to determine the exact distribution of Z_n. Therefore, we adopt a Monte Carlo simulation to further determine the control limits. The above process description also reflects the online monitoring of our methods, that is, the in-control state is determined by using the data in Phase I, and the risk-adjusted model is generated by training the historical data; then, the surgical risks are monitored consecutively on each of cases in Phase II. Thus the proposed method can online determine when the underlying risk of surgery deviates to the historical average level (in-control state).

3. Results

3.1. Experimental protocol

To explore the monitoring performance of the proposed methods, we run a series of simulations using simulated data and a set of real data experiments. During the experiment, we compare the four methods we have proposed, and also introduce ESOP and CUSUM as comparison methods. The reason for choosing only ESOP is that the literature [18] have indicated that ESOP can detect changes in location parameters and scale parameters, which is more efficient than other existing methods in detecting the small shifts. Similarly to the previous study [18], the CUSUM control chart [20], which is another commonly used monitoring method, is also selected in the comparison. It is worth mentioning that ESOP is a two-sided chart that can be used to monitor upward shifts and downward shifts, but in this section we only consider the case where surgical performance is deterioration using a one-sided chart--upper-sided chart ($p_n{}^* > p_n$). This is because the principle of monitoring improvement in surgical performance is the same as deterioration, and monitoring deterioration makes more sense in reality.

The performance of each control chart is assessed in terms of average run length (ARL), which is defined as the average of the run length distribution. Specifically, run length is a random variable equal to the number of samples required to observe the first out-of-control signal [28]. We select ARL and standard deviation of ARL as specific evaluation indicators, which are expressed as ARL and standard deviation of ARL respectively. As indicators for evaluating the monitoring methods of the surgical outcomes, the smaller values of ARL and standard deviation of ARL, the better the corresponding method.

Now we make some explanations on the setting of experimental parameters. Control limit (CL) is determined by the in control (IC) ARL, considering comprehensively, we set IC ARL in the simulation data experiment as 400, and set it to 500 in the real-world data experiment. Similar to ESOP, we assign values 0.5 and -1.386 to β and α respectively. In practice, β and α can be estimated according to an IC sample. The a value of the time-varying control limits is set to 0.014, which is derived from $MFIR_{adj}$ = 0.99 at t = 400. λ's choice is very similar to choose the smoothing parameter of a traditional EWM-type chart. Liu et al. suggest choosing $\lambda \in [0.005, 0.1]$ on the basis of empirical results, and verify that a small value of λ is good at quickly detecting small shifts, while a large value of λ is good at quickly detecting large shifts [18]. Since our purpose is not to verify the effect of λ's value on the monitoring performance, we take an intermediate value 0.01 as the value of λ. According to [18], we set $\rho = 2$ in CUSUM, which ensures it has better performance. We consider from two aspects: the fixed shifts and the random effects, which are regulated by parameters δ and τ, respectively. In order to make the experiment results credible, all the ARL results are obtained from 10,000 replications.

3.2. Simulation performance evaluation

First, we use ESOP, CUSUM, SESOP and STSSO charts to detect the deterioration in surgical performance. The results are shown in Tables 1 and 2, where Table 1 corresponds to fixed shifts and Table 2 records the results under random effects. In Table 1, we can see that when δ is less than 1.0, that is, the fixed drift is small, both SESOP and STSSO perform better than ESOP and CUSUM in different aspects. Specifically, SESOP has always been in a dominant position in ARL, indicating that SESOP can always detect shifts in the process earlier. In terms of standard deviation of ARL, STSSO shows better stability than other three methods when shifts are small and SESOP has smallest standard deviation when δ is between 0.4 and 1.0. Therefore, when δ is smaller than 1.0, if users focus on the stability of control chart, they can use STSSO when the study aims to detect small shifts, and adopt SESOP while the expected shifts are large. The CUSUM chart performs better when the fixed shifts are large. Specifically, when $\delta \geq 1.0$, the CUSUM is the best one in terms of ARL. Table 2 reflects a different situation from Table 1. When τ is less than 0.5, CUSUM has a smaller ARL. But with random effects gradually increasing, ESOP and SESOP outperform CUSUM in ARL. On the other hand, STSSO is still in absolute advantage in terms of ARL standard deviation, which means STSSO is most stable when random effects occur. We translate Table 1 and 2 into Figure 2 for easy demonstration of the difference.

Table 1. ARL performance of ESOP, CUSUM, SESOP and STSSO with different fixed shifts.

$logit(p_n)$		$x_n\beta + \alpha + \delta$										
δ		0.0	0.1	0.2	0.3	0.4	0.5	0.7	1.0	1.5	2.0	3.0
ESOP	mean	400.00	225.84	141.68	99.65	72.96	56.60	38.17	24.43	14.51	10.30	6.72
	sd	424.94	225.67	133.10	87.20	60.35	44.53	27.83	16.07	8.91	5.90	3.60
CUSUM	mean	400.00	262.27	169.32	117.48	86.64	62.96	39.12	23.40*	13.00*	8.94*	5.76*
	sd	402.97	258.00	162.71	107.51	77.53	53.70	30.21	16.03	7.55*	4.51*	2.32*
SESOP	mean	400.00	215.46*	138.32*	95.16*	70.88*	54.21*	36.82*	23.43	14.27	9.99	6.70
	sd	421.18	214.57	132.87	82.86	58.28*	41.95*	26.09*	15.40*	8.58	5.65	3.54
STSSO	mean	400.00	283.07	210.94	167.36	137.76	115.76	86.23	61.23	39.84	29.43	20.19
	sd	260.66*	166.65*	110.21*	79.02*	60.49	46.80	31.96	20.29	11.92	8.30	5.08

Note: * is used to mark the optimal value.

Table 2. ARL performance of ESOP, CUSUM, SESOP and STSSO with random effects.

$logit(p_n)$		$x_n\beta + \alpha + \tau\vartheta$					
τ		0.0	0.1	0.2	0.3	0.4	0.5
ESOP	mean	400.00	396.74	370.14	341.15	306.43	264.83
	sd	424.94	422.55	395.64	357.63	316.09	267.05
CUSUM	mean	400.00	366.78*	320.08*	284.71*	263.74*	242.69*
	sd	402.97	367.54	314.94	277.81	256.87	237.59
SESOP	mean	400.00	383.00	360.47	337.48	295.62	261.60
	sd	421.18	402.07	381.69	347.33	304.14	266.71
STSSO	mean	400.00	395.52	378.51	360.46	335.95	304.93
	sd	260.66*	264.48*	248.54*	230.55*	206.25*	186.32*
$logit(p_n)$		$x_n\beta + \alpha + \tau\vartheta$					
τ		0.0	0.7	1.0	1.5	2.0	3.0
ESOP	mean	400.00	195.86	125.49	72.00*	50.18*	32.84*
	sd	424.94	194.44	114.97	60.13	39.45*	23.20*
CUSUM	mean	400.00	202.00	165.46	125.40	101.44	73.18
	sd	402.97	195.64	157.07	117.76	92.64	64.60
SESOP	mean	400.00	191.21*	125.16*	72.84	50.92	33.45
	sd	421.18	192.16	118.33	61.15	39.94	23.79
STSSO	mean	400.00	252.40	191.81	131.13	99.99	72.91
	sd	260.66*	142.37*	97.33*	56.76*	40.21	26.89

Note: * is used to mark the optimal value.

Then, in order to verify the improvement of surgical performance monitoring by time-varying control limits, we select SESOP and SESOP-MFIR for the comparative experiment, and change f value to further explore the effect of different f values on the performance. It is clear that changing the asymptotic control limit to the time-varying control limit can improve the monitoring performance, especially about ARL. Table 3 and 4 show that the dynamic control limit has a greater monitoring ability than the asymptotic control limit, and the smaller the f value is, the earlier the shifts can be detected. Specifically, when shifts are fixed values, SESOP-MFIR with f = 0.3 outperforms SESOP-MFIR with f = 0.5 in terms of the ARL, and SESOP-MFIR with f = 0.5 is more efficient than SESOP in detecting the fixed shifts. Results indicate that the addition of time-varying control limit can detect the early-stage anomalies more quickly. The value of f affects the trend of the control limit at the initial stage, the smaller the value, the more sensitive it is to drift at the initial stage. But it should be noted that too small f value may lead to a few unstable results. The time-varying control limit increases the standard deviation of ARL in the case of small shifts. However, the f value influences the standard deviation in a different way when shifts are large. Therefore, it is reasonable to believe that if we only focus on the optimization of the ARL, we can continue to reduce the value of f to achieve this goal. A visual display of the results is shown in Figure 3.

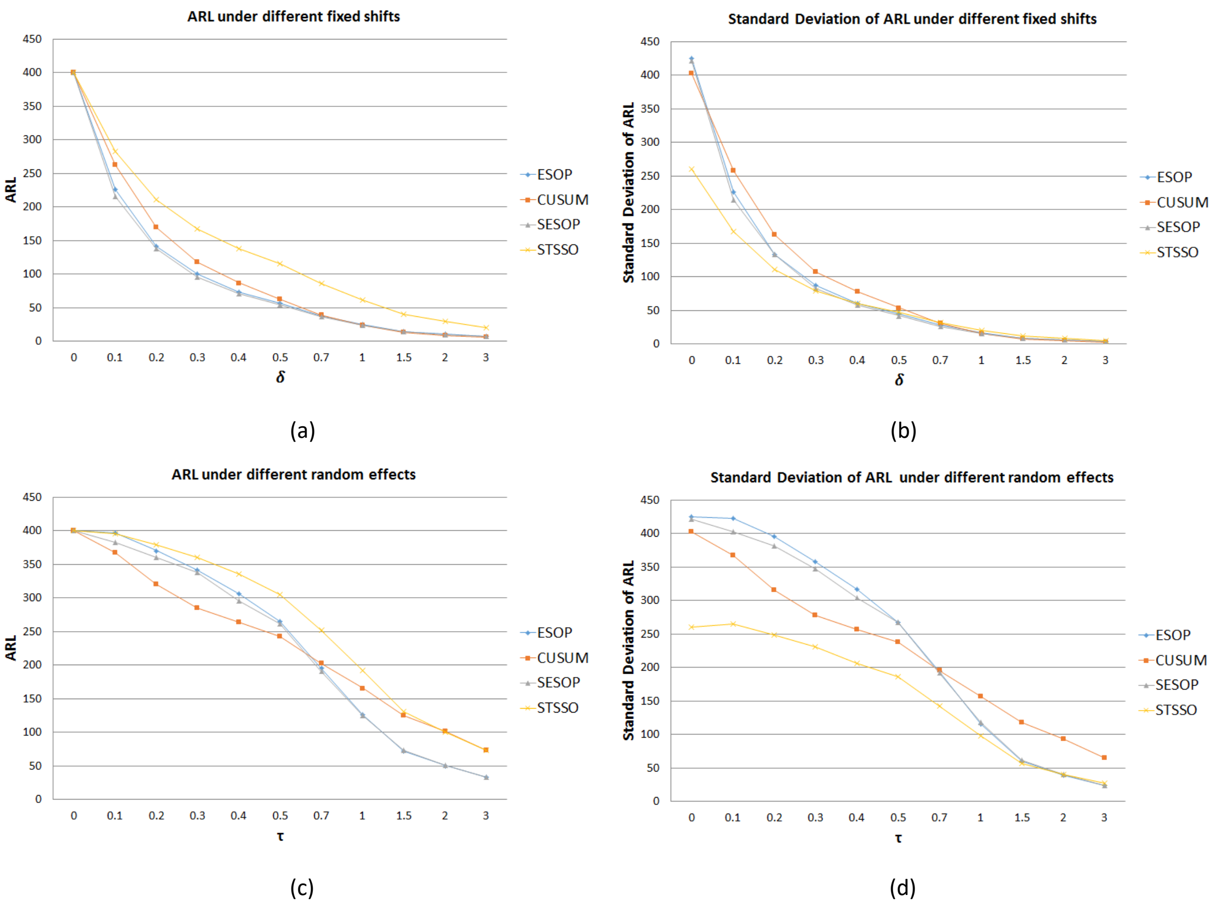

Figure 2. ARL performance of ESOP, CUSUM, SESOP and STSSO with different fixed shifts (a, b) and random effects (c, d).

Table 3. ARL performance of SESOP and SESOP-MFIR with different fixed shifts.

$logit(p_n)$		$x_n\beta+\alpha+\delta$					
δ		0.0	0.1	0.2	0.3	0.4	0.5
SESOP	mean	400.00	225.84	141.68	99.65	72.96	56.60
	sd	424.94*	225.67	133.10	87.20	60.35	44.53
SESOP-MFIR (f = 0.5)	mean	400.00	207.47	130.35	87.84	64.43	49.46
	sd	452.21	224.02*	129.92	83.21	57.09	42.34
SESOP-MFIR (f = 0.3)	mean	400.00	203.81*	113.71*	77.46*	56.68*	44.13*
	sd	512.76	246.41	126.16*	76.64*	53.30*	38.51*
$logit(p_n)$		$x_n\beta+\alpha+\delta$					
δ		0.0	0.7	1.0	1.5	2.0	3.0
SESOP	mean	400.00	38.17	24.43	14.51	10.30	6.72
	sd	424.94*	27.83	16.07	8.91	5.90	3.60
SESOP-MFIR (f = 0.5)	mean	400.00	33.15	21.18	12.62	8.70	5.69
	sd	452.21	25.79	15.34	8.06	5.24	3.22
SESOP-MFIR (f = 0.3)	mean	400.00	28.75*	18.66*	11.17*	7.82*	5.08*
	sd	512.76	23.46*	13.83*	7.41*	4.92*	2.99*

Note: * is used to mark the optimal value.

Table 4. ARL performance of SESOP and SESOP-MFIR with random effects.

$logit(p_n)$		$x_n\beta+\alpha+\tau\vartheta$					
τ		0.0	0.1	0.2	0.3	0.4	0.5
SESOP	mean	400.00	383.00*	360.47*	337.48	295.62	261.60
	sd	421.18*	402.07*	381.69*	347.33*	304.14*	266.71*
SESOP-MFIR (f = 0.5)	mean	400.00	384.16	368.83	343.95	293.34	252.15
	sd	452.21	421.30	404.92	387.79	317.97	273.59
SESOP-MFIR (f = 0.3)	mean	400.00	394.39	373.87	326.67*	284.80*	243.87*
	sd	512.76	511.73	480.81	413.42	367.05	309.01
$logit(p_n)$		$x_n\beta+\alpha+\tau\vartheta$					
τ		0.0	0.7	1.0	1.5	2.0	3.0
SESOP	mean	400.00	191.21	125.16	72.84	50.92	33.45
	sd	421.18*	192.16*	118.33	61.15	39.94	23.79
SESOP-MFIR (f = 0.5)	mean	400.00	189.75	121.50	67.01	46.87	29.99
	sd	452.21	199.78	122.15	59.84	39.22	23.18
SESOP-MFIR (f = 0.3)	mean	400.00	176.97*	105.74*	58.32*	40.59*	26.87*
	sd	512.76	216.39	116.37*	55.52*	34.89*	22.32*

Note: * is used to mark the optimal value.

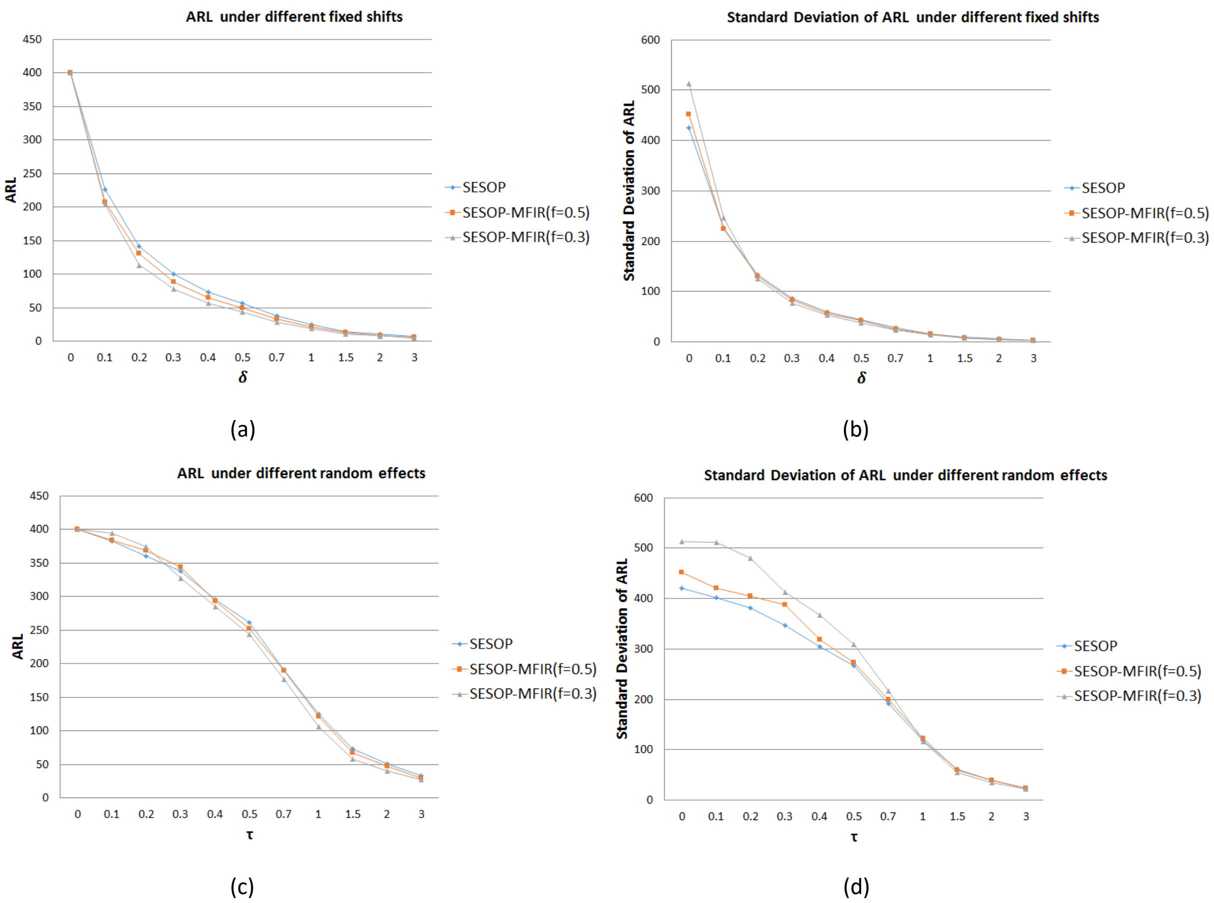

Figure 3. ARL performance of SESOP and SESOP-MFIR with different fixed shifts (a, b) and random effects (c, d).

Finally, we continue the previous experiments, reducing the f value, and do a comparative experiment on ESOP, CUSUM, SESOP-MFIR with f = 0.1 and STSSO-MFIR with f = 0.2. Since the influence of f value on ARL and the standard deviation of ARL has been verified, in consideration of memory consumption and time cost, we focus on ARL in this experiment. As shown in Table 5, SESOP-MFIR and STSSO-MFIR are both more effective in detecting shifts than ESOP and CUSUM in most cases, especially the advantages of STSSO-MFIR are more obvious. For example, when shift δ = 0.1, from CUSUM to SESOP-MFIR, the value of ARL is reduced from 262.27 to 135.49, and finally optimized to 77.91 of STSSO-MFIR. The difference value between STSSO-MFIR and CUSUM is as high as 184.36, which is a huge improvement. Another point to note is that, according to the previous experimental results, the smaller the value of f, the better the result of ARL. It could be observed that STSSO-MFIR with f = 0.2 has lower ARL results than SESOP-MFIR with f = 0.1, suggesting that STSSO-MFIR with f = 0.2 may be more efficient in detecting changes in practical implementation. The visualization results are shown in Figure 4.

Table 5. The performance of ARL of ESOP, CUSUM, SESOP-MFIR (f = 0.1) and STSSO-MFIR (f = 0.2).

$logit(p_n)$	$x_n\beta+\alpha+\delta$										
δ	0.0	0.1	0.2	0.3	0.4	0.5	0.7	1.0	1.5	2.0	3.0
ESOP	400.00	225.84	141.68	99.65	72.96	56.60	38.17	24.43	14.51	10.30	6.72
CUSUM	400.00	262.27	169.32	117.48	86.64	62.96	39.12	23.40	13.00	8.94	5.76
SESOP-MFIR	400.00	135.49	78.60	57.65	43.89	33.82	23.37	16.02	9.51	6.78	4.49
STSSO-MFIR	400.00	77.91*	41.05*	31.11*	25.81*	22.34*	16.53*	11.79*	7.77*	5.69*	3.69*
$logit(p_n)$	$x_n\beta+\alpha+\tau\vartheta$										
τ	0.0	0.1	0.2	0.3	0.4	0.5	0.7	1.0	1.5	2.0	3.0
ESOP	400.00	396.74	370.14	341.15	306.43	264.83	195.86	125.49	72.00	50.18	32.84
CUSUM	400.00	366.78*	320.08*	284.71	263.74	242.69	202.00	165.46	125.40	101.44	73.18
SESOP-MFIR	400.00	387.94	325.73	263.50	227.76	183.66	115.57	75.99	44.31	32.71	20.89
STSSO-MFIR	400.00	367.67	324.12	244.24*	173.75*	102.46*	56.55*	36.29*	24.99*	18.95*	14.57*

Note: * is used to mark the optimal value.

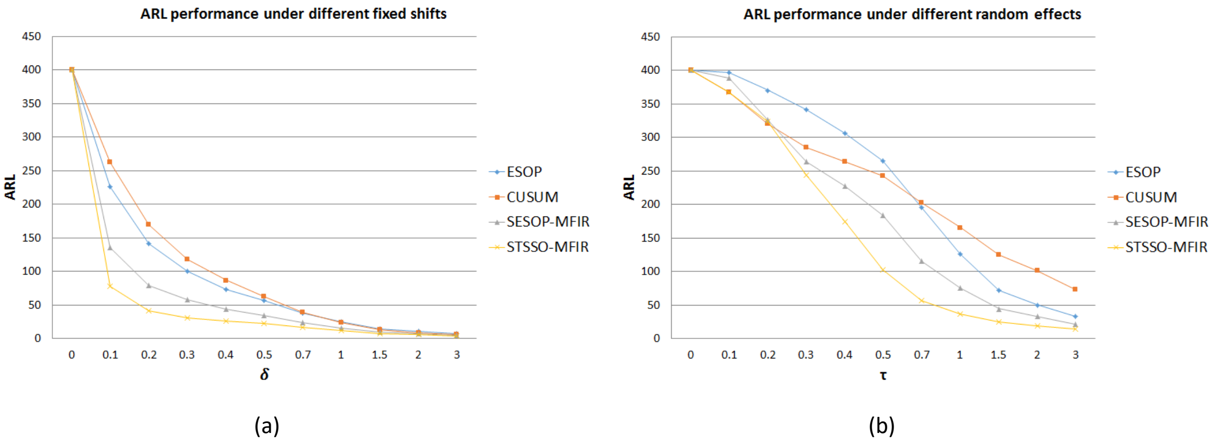

Figure 4. ARL performance of ESOP, CUSUM, SESOP-MFIR and STSSO-MFIR with different fixed shifts (a) and random effects (b).

3.3. A practical application in Hong Kong

In this section, we apply ESOP, SESOP, STSSO, SESOP-MFIR, STSSO-MFIR and CUSUM to the SOMIP program from the Hospital Authority of Hong Kong. The data source is the surgical outcomes of hospital A between 2009 and 2013, We follow the rule as Liu et al. [18] to estimate the risks of 30-day mortality based on surgical outcomes from 2009 to 2013. Specifically, data from 2009 to 2012 utilizes $p_n{}^* = \frac{exp\,(x_n\beta+\alpha+\omega)}{1+exp(x_n\beta+\alpha+\omega)}$, then we calculate the charting statistics Z_n for each data in 2013 according to the principle of each method. Starting from the first case in 2013, we can use the established method to online monitor the risk at each surgery case. In order to obtain the control limits, the bootstrapping technology is adopted to process the data from 2009 to 2012, and ARL is set as 500, which is because 500 basically matched the annual operation volume of a hospital with the same size. Similar to the previous simulation, we still only show the results of monitoring deterioration in this section.

ESOP, SESOP, STSSO, SESOP-MFIR, STSSO-MFIR and CUSUM evaluate hospital A's performance in 2013 based on the overall performance from 2009 to 2012, specifically, we fit surgical data from 2009 to 2012 with a logical model, then use this model to monitor adjusted risk of the data in 2013. The results are shown in Figure 5. In Figure 5(a), after the 320th case, several deteriorations are detected by ESOP. If the SESOP is used (Figure 5(b)), the deterioration, which is not detected, may occurred at early stage. Similarly, neither STSSO nor CUSUM detects the possible deteriorations. In Figure 5(d) and 5(e), we can visually observe that the control limits are time-varying, and SESOP-MFIR discovers volatilities after 300th case, in addition, both SESOP-MFIR and STSSO-MFIR can detect the deteriorations between 100th case and 170th case, which means that the time-varying control limits make the initial small shift more easily detected, indicating that SESOP-MFIR and STSSO-MFIR are more sensitive to small changes at early stage, it also corresponds to the results in the simulation experiments.

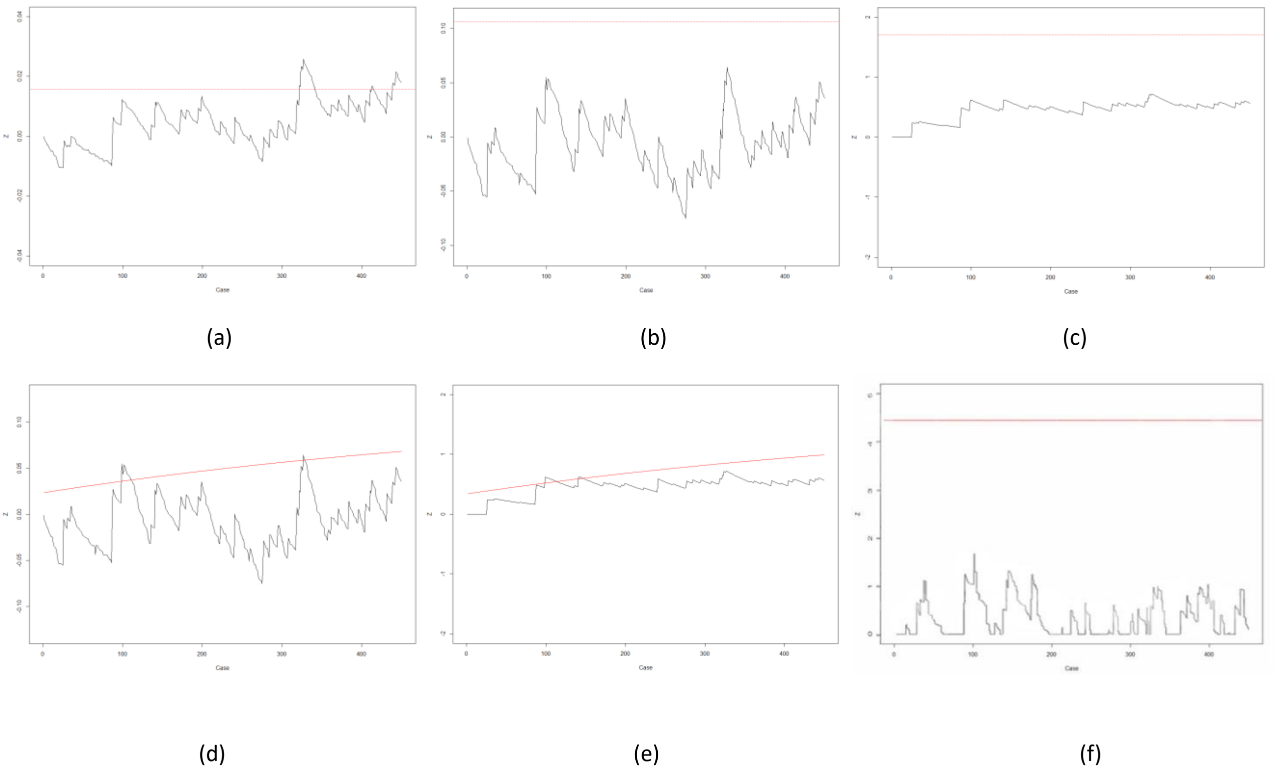

Figure 5. Monitoring charts for hospital A: (a) ESOP chart; (b) SESOP chart; (c) STSSO chart; (d) SESOP-MFIR chart; (e) STSSO-MFIR; (f) CUSUM chart. The X-axis represents the number of cases, the Y-axis represents the Z-statistic, and the red dotted line represents the control limit of the corresponding method. This graph reflects the changes in the Z-statistic with the addition of cases and the sensitivity of the control limit of different methods to monitoring for deterioration.

4. Discussion

Based on the several sets of simulations and a real-world application in Hong Kong, SESOP, STSSO, SESOP-MFIR and STSSO-MFIR can be improved in different degrees on surgical outcome monitoring performance compared with ESOP.

The first finding worth noting is that standardizing variables can effectively reduce the ARL when an out-of-control (OC) signal is sent, that is, it can react to changes in a timely manner. The one-sided test statistics T can bring significant optimization in terms of ARL standard deviation, especially when the shifts are small, which shows that STSSO has more stable performance during the monitoring process.

Subsequently, we combine a kind of fast initial response (FIR) feature with SESOP and STSSO respectively, which named SESOP-MFIR and STSSO-MFIR, to transform the asymptotic control limits into the time-varying control limits. Through the experimental results, we observe that the dynamic control limits can greatly reduce the ARL, which brings more rapid responses to the monitoring process, and is great significance for postoperative risk monitoring. It's a slight pity that the dynamic control limits perform slightly worse than the asymptotic control limits in the IC state (shift is zero).

Therefore, we recommend using SESOP or STSSO at the beginning of the process. If there is a change, it will be converted to SESOP-MFIR or STSSO-MFIR. This combination can bring satisfactory monitoring results. Of course, users can develop monitoring plans according to their own needs and the characteristics of methods.

5. Conclusion

In this article, we propose EWMA control charting methods--SESOP and STSSO, then add the FIR features based on these two, and further put forward SESOP-MFIR and STSSO-MFIR. The main difference between them is the control limit, the former is asymptotic, the latter is time-varying. Both the location and scale parameters can be monitor by these new charting methods in the surgical outcomes model simultaneously, and they have their own expertise, which can be confirmed in simulation experiments compared to the ESOP and CUSUM. The four methods improve the performance of surgical outcomes monitoring in different degrees and aspects compared to those existing methods. Specifically, SESOP has more excellent and stable performance in the optimization of ARL, which means it can detect a shift faster, whether the shift is large or small. Correspondingly, STSSO combines a score test statistics, so that it has better performance of ARL standard deviation when small volatilities occur, that is, monitoring in this case is more stable. Due to the time-varying control limits, SESOP-MFIR and STSSO-MFIR have significantly improvement on efficiency of shifts detection. Subsequently, we apply the proposed monitoring methods to the hospital A's data from SOMIP project of the Hong Kong's Hospital Authority and used logistic regression models to fit the binary surgical outcomes. It can be visually observed that the control limits of SESOP-MFIR and STSSO-MFIR are time-varying, and they can detect early shifts that other methods cannot easily find, which means these two control charts are more sensitive to small changes at early stage, it also corresponds to the results in the simulation experiments. The comprehensive results of the real-world application show that the performance of hospital A in 2013 is more excellence than that of the previous three years. In general, we optimize the surgical outcomes monitoring control charts through the standardization of variables, the score test statistic and the combination of FIR features, both the simulation study and the practical application verified their excellence performance. By using the risk adjusted model that is built in Phase I, the proposed method can online monitor the surgical risks, which has great significance for making timely alters.

Availability of data and materials

The datasets used and analyzed during the current study are available from the SOMIP upon reasonable requesting procedure.

Funding

The work is supported by grants from National Natural Science Foundation of China (No. 71872146 and No. 31701150).

Authors' contributions

Jiaqi Liu leaded the method application, experiment conduction and the result analysis. Paul B.S.

Lai participated in the data extraction and preprocessing. Jiayin Wang participated in the manuscript revision. Xin Lai provided theoretical guidance and the revision of this paper.

Conflict of interest

All authors declare no conflicts of interest in this paper.

References

1. W. H. Woodall, S. L. Fogel, S. H. Steiner, The monitoring and improvement of surgical outcome quality, *J. Qual. Technol.*, **47** (2015), 383–399.
2. K. Paynabar, J. Jin, A. B. Yeh, Phase I risk-adjusted control charts for monitoring surgical performance by considering categorical covariates, *J. Qual. Technol.*, **44** (2012), 39–53.
3. M. Douglas, Z. Song, Y. Liu, J. Zhang, A comparative study of memory-type control charts based on robust scale estimators, *Qual. Reliab. Eng. Int.*, **34** (2018), 1079–1102.
4. J. S. Hunter, The exponentially weighted moving average, *J. Qual. Technol*, **18** (1986), 203–210.
5. A. Ingolfsson, E. Sachs, Stability and sensitivity of an EWMA controller, *J. Qual. Technol.*, **25** (1993), 271–287.
6. S. V. Crowder, M. D. Hamilton, An EWMA for monitoring a process standard deviation, *J. Qual. Technol.*, **24** (1992), 12–21.
7. R. B. Crosier, A new two-sided cumulative sum quality control scheme, *Technometrics*, **28** (1986), 187–194.
8. L. Shu, W. Jiang, K. L. Tsui, A comparison of weighted CUSUM procedures that account for monotone changes in population size, *Stat. Med.*, **30** (2011), 725–741.
9. J. J. Pignatiello, G. C. Runger, Comparisons of multivariate CUSUM charts, *J. Qual. Technol.*, **22** (1990).
10. J. Lovegrove, O. Valencia, T. Treasure, C. Sherlaw-Johnson, S. Gallivan, Monitoring the results of cardiac surgery by variable life-adjusted display, *Lancet,* **350** (1997), 1128–1130.
11. T. Treasure, O. Valencia, C. Sherlaw-Johnson, S. Gallivan, Surgical performance measurement, *Health Care Manag. Sci.*, **5** (2002), 243–248.
12. C. W. Champ, W. H. Woodall, Signal probabilities for runs supplementing a shewhart control chart, *Commun. Statist. Simulat. Comput.*, **26** (1997), 1347–1360.
13. J. M. Lucas, Combined Shewart-CUSUM quality control schemes, *J. Qual. Technol.*, **14** (1982), 51–59.
14. L. S. Nelson, The Shewhart control chart-tests for special causes, *J. Qual. Technol.*, **16** (1984), 237–239.
15. J. Neuburger, K. Walker, C. Sherlaw-Johnson, J. Meulen, D. A. Cromwell, Comparison of control charts for monitoring clinical performance using binary data, *BMJ Qual. Safety*, **26** (2017), 919–928.
16. S. H. Steiner, W. H. Woodall, Debate: what is the best method to monitor surgical performance, *BMC Surg.*, **16** (2016), 15.
17. X. Zhang, W. H. Woodall, Reduction of the effect of estimation error on in-control performance for risk-adjusted Bernoulli CUSUM chart with dynamic probability control limits, *Qual. Reliab. Eng. Int.*, **33** (2016), 381–386.

18. L. Liu, X. Lai, J. Zhang, F. Tsung, Online profile monitoring for surgical outcomes using a weighted score test, *J. Qual. Technol.*, **1** (2018), 88–97.
19. D. A. Cook, S. H. Steiner, R. J. Cook, V. T. Farewell, A. P. Morton, Monitoring the evolutionary process of quality: risk-adjusted charting to track outcomes in intensive care, *Crit. Care. Med.*, **31** (2003), 1676–1682.
20. S. H. Steiner, R. J. Cook, V. T. Farewell, T. Treasure, Monitoring surgical performance using risk adjusted cumulative sum charts, *Biostatistics*, **1** (2000), 441–452.
21. J. Yue, X. Lai, L. Liu, P. Lai, A new VLAD-based control chart for detecting surgical outcomes, *Stat. Med.*, **36** (2017), 4540–4547.
22. X. Zhang, W. H. Woodall, Dynamic probability control limits for risk-adjusted Bernoulli CUSUM charts, *Stat. Med.*, **34** (2015), 3336–3348.
23. D. A. Cook, M. Coory, R. A. Webster, Exponentially weighted moving average charts to compare observed and expected values for monitoring risk-adjusted hospital indicators, *BMJ Qual. Safety*, **20** (2011), 469–474.
24. T. Treasure, S. Gallivan, C. Sherlaw-Johnson, Monitoring cardiac surgical performance: A commentary, *J. Thorac. Cardiov. Sur.*, **128** (2004), 823–825.
25. O. Grigg, V. T. Farewell, A risk-adjusted sets method for monitoring adverse medical outcomes, *Stat. Med.*, **23** (2004), 1593–1602.
26. S. H. Steiner, EWMA control charts with time-varying control limits and fast initial response, *J. Qual. Technol.*, **31** (1999), 75–86.
27. J. M. Lucas, R. B. Crosier, Fast initial response for CUSUM quality-control schemes: give your CUSUM a head start, *Technometrics*, **24** (1982), 199–205.
28. A. Haq, J. Brown, E. Moltchanova, Improved fast initial response features for exponentially weighted moving average and cumulative sum control charts, *Qual. Reliab. Eng. Int.*, **30** (2014), 697–710.
29. S. Knoth, Fast initial response features for EWMA control charts, *Stat. Pap.*, **46** (2005), 47–64.
30. M. J. Silvapulle, P. Silvapulle, A score test against one-sided alternatives, *J. Am. Stat. Assoc.*, **90** (1995), 342–349.
31. S. K. Sinha, Bootstrap tests for variance components in generalized linear mixed models, *Can. J. Stat.*, **37** (2009), 219–234.