Computers & Industrial Engineering 160 (2021) 107568

Contents lists available at ScienceDirect

Computers & Industrial Engineering

journal homepage: www.elsevier.com/locate/caie

A risk-adjusted approach to monitoring surgery for survival outcomes based on a weighted score test

Xin Lai [a], Xiao Li [a], Liu Liu [b], Fugee Tsung [c,*], Paul B.S. Lai [d], Jiayin Wang [a], Xuanping Zhang [a], Xiaoyan Zhu [a], Jiaqi Liu [a]

[a] School of Computer Science and Technology, Xi'an Jiaotong University, Xi'an, China
[b] School of Mathematics and VC & VR Lab, Sichuan Normal University, Chengdu, Sichuan, China
[c] Department of Industrial Engineering and Decision Analytics, Hong Kong University of Science and Technology, Kowloon, Hong Kong
[d] Department of Surgery, The Chinese University of Hong Kong, Hong Kong

ARTICLE INFO

Keywords:
Cox model
EWMA chart
Risk adjustment
Survival time
Score test statistics

ABSTRACT

In programs monitoring surgical quality, risk-adjusted control charts have been used widely to detect changes in surgical performance. Ignoring the survival time may lead to information loss and thus attenuate the monitoring efficiency. However, previous methods based on survival time information only focus on detecting the change in average risk. The stability of surgical performance measured by scale parameters could be of interest in monitoring. In this study, we extend the risk-adjusted monitoring approach to include survival outcomes, in which both the location and scale parameters are monitored simultaneously. Based on the weighted score test for the Cox model, we propose to use an exponentially weighted moving average chart to monitor changes in average surgical risk and the existence of its variance, which could be of interest in practical surgical monitoring programs. Simulation results indicate that the proposed method detects changes in the variance of surgical performance and small shifts in surgical risk more efficiently than existing cumulative sum methods. In addition, the proposed method shows good efficiency for various magnitudes of shift. The proposed chart was applied to a data set from the Surgical Outcome Monitoring and Improvement Program in Hong Kong, identifying an improvement in a hospital's outcomes.

1. Introduction

Statistical process control (SPC) methods (Shang, Tsung, & Zou, 2013; Shen, Zou, Jiang, & Tsung, 2013; Castagliola & Tsung, 2005; Shu & Tsung, 2003; Chen et al., 2007; Jiang and Tsui, 2008) have been adapted to monitor surgical risks and to help identify the root cause of problems (Paynabar et al., 2012; Steiner et al., 2000; Cook et al., 2011; Steiner, 2014). The SPC methods have gained popularity since Treasure, Taylor, and Black (1997) and Waldie (1998) pioneered the use of the control charts to monitor surgical performance at Bristol Hospital because of their advantages in detecting changes in unobserved states.

A risk-adjusted model with surgical factors was employed to accommodate the heterogeneity of patient-level risks in surgery, and then charting statistics were adopted to monitor the adjusted risks (Woodall et al., 2015). Specifically, Steiner et al. (2000) proposed a risk-adjusted cumulative sum (RA CUSUM) chart to detect a deterioration in 30-day mortality, and Cook et al. (2011) used the Exponential Weighted Moving Average method to monitor adjusted surgical risks (RA EWMA). To monitor the variation of surgical risks, which is an important dimension when evaluating surgical performance, Liu et al. (2018) proposed an online weighted score test to detect changes in both average risk and variance.

The Hospital Authority (HA) of Hong Kong launched the Surgical Outcome Monitoring Improvement Program (SOMIP) in 2008, where the surgical risks in each public hospital were evaluated annually. The Variable Life Adjusted Display (VALD) (Lovegrove et al., 1997; Sherlaw-Johnson, 2005; Grigg and Farewell 2004) was employed to plot changes in the adjusted surgical risks each hospital by using logistic regression to fit the outcome and patient-level factors. Based on that, the CUSUM (Steiner et al., 2000) or EWMA (Grigg and Spiegelhalter, 2007; Cook et al., 2008; Yue et al., 2017) method was used to detect significant deteriorations in surgical performance. More VLAD results in SOMIP can

* Corresponding author.
E-mail addresses: laixin@xjtu.edu.cn (X. Lai), lixiao19980319@stu.xjtu.edu.cn (X. Li), liuliu@sicnu.edu.cn (L. Liu), season@ust.hk (F. Tsung), paullai@surgery.cuhk.edu.hk (P.B.S. Lai), zxp@mail.xjtu.edu.cn (X. Zhang), zhu.xy@mail.xjtu.edu.cn (X. Zhu), liujiaqi@stu.xjtu.edu.cn (J. Liu).

https://doi.org/10.1016/j.cie.2021.107568
Received 17 February 2021; Received in revised form 16 July 2021; Accepted 19 July 2021
Available online 22 July 2021
0360-8352/© 2021 Elsevier Ltd. All rights reserved.

be found in Yuen (2013).

In the above approaches and programs, 30-day mortality is the most commonly used outcome for evaluation because it is easy to interpret and can provide a good measure for tracking improvements in quality (Mant, 2001; Merkow et al., 2013; In et al., 2016). However, as advances in life support systems prolong survival, there are increasing concerns that postoperative death could be artificially delayed to beyond 30 days if 30-day mortality is the only standard used for quality assessment (Walters et al., 2014; In et al., 2016). Even though some scholars have proposed more extended periods to mitigate the effect of 30-day mortality, such as 90-day mortality (Talsma et al., 2014; Mise et al., 2015), the use of a binary outcome with a fixed cut-off time is still likely to lose important information (Steiner and Jones, 2010). For example, patients who died on the first day of a study period may have experienced a different quality of surgery to those with the same preoperative risks who died later. Still, methods based on a binary outcome will give the same evaluation and monitoring results.

Sego et al. (2009) considered the log-logistic and Weibull distribution for the baseline hazard in monitoring, and the log-likelihood ratio score was derived to update the CUSUM statistic. Following the same parametric setting for baseline hazard and likelihood ratio, Steiner and Jones (2010) proposed an EWMA method to monitor continuous survival time. Although those methods are applicable when monitoring survival outcomes, the parametric assumption of baseline hazard may not be correct in practice in some scenarios and could then bias the charting statistic. Biswas and Kalbfleisch (2008) and Gandy et al. (2010) considered the Cox proportional hazard model for risk-adjusted CUSUM. The baseline hazard was unspecified and could be estimated from a large population sample. Recently, Grigg (2019) developed a survival time risk-adjusted N-division (STRAND) chart for monitoring survival outcomes online. He divided the continuous survival time into N strands and detected the change in the odds ratio of each strand by applying the RA EWMA described in Grigg and Spiegelhalter (2007). In addition, Gan et al. (2020) and Keefe et al. (2017) accommodated the information in survival time by daily updating the RA-CUSUM. However, those methods were designed to detect changes in the mean value of the surgical risk. Therefore, variance, which reflects the stability of surgical performance (Liu et al., 2018), may not be detected efficiently by these survival monitoring approaches. Therefore, it would be helpful to develop a risk-adjusted survival time control chart with unspecified baseline hazards that monitor both the mean (location parameter) and variance levels (scale parameter) and may evaluate surgical performance more comprehensively.

This paper proposes a control chart for continuous survival time after surgery, aiming to detect changes in the average surgical risk and its stability. By adapting the score test for random effects, as in Liang (1987), Zhu and Zhang (2006), and Liu et al. (2018), changes in the average hazard and its variance can be detected. In particular, the baseline hazard function can be estimated nonparametrically, and an EWMA procedure is applied to update the charting statistic after each surgical outcome sequentially.

In the next section, the proposed control chart, constructed from the score test and the EWMA updating scheme, is introduced. A simulation study is presented in Section 3 to assess our method's performance under different process changes. In Section 4, the time-to-death data from SOMIP in Hong Kong illustrate the applicability of the proposed chart. Finally, further discussion is presented in Section 5.

2. Risk-Adjusted control chart for survival outcome

Suppose the underlying death time of the ith patient following surgery is T_i^*, and the follow-up time is C_i, then the observed time $T_i = \min(T_i^*, C_i)$. The censoring indicator is defined as $\delta_i = 1$ if the ith patient died after surgery (i.e., $T_i^* \le C_i$) and $\delta_i = 0$ if the real time of death T_i^* is beyond the study time (i.e., $T_i^* > C_i$). Given the covariate x_i, the hazard function of the ith patient at time t can defined as

$$h(t|x_i) = \lim_{\Delta t \to 0} \frac{\Pr\left(t \le T_i^* < t + \Delta t \middle| T_i^* \ge t, x_i\right)}{\Delta t} \tag{1}$$

We further assume the proportional hazard model (Cox, 1972)

$$h(t|x_i) = h_0(t) e^{x_i\beta + \gamma} \tag{2}$$

where the baseline hazard is unspecified and considered only related to t. The measured risk factor x_i exerts the effect through an exponential function. A fixed γ can describe the change in the average of surgical performance (Biswas and Kalbfleisch, 2008; Gandy et al., 2010). Specifically, $\gamma = 0$ indicates that performance is normal (i.e., in-control state) and $\gamma > 0$ indicates a deterioration in surgical performance (i.e., out-of-control state) because the hazard increases to a level higher than the normal state. In practice, the clinicians are interested in detecting when surgical performance starts to deteriorate ($\gamma > 0$). It may be meaningful for regulators to monitor improvements ($\gamma < 0$). Therefore, the process monitoring is equivalent to sequentially testing

$$H_0 : \gamma = 0 vs. H_1 : \gamma \ne 0 \tag{3}$$

For the ith case, the likelihood ratio between the alternative and the null hypotheses is

$$LR_i = \frac{L_i(\gamma)}{L_i(0)} = \frac{[f(t_i|x_i,\gamma)]^{\delta_i}[S(t_i|x_i,\gamma)]^{1-\delta_i}}{[f(t_i|x_i,0)]^{\delta_i}[S(t_i|x_i,0)]^{1-\delta_i}} = \frac{[h(t_i|x_i,\gamma)]^{\delta_i} S(t_i|x_i,\gamma)}{[h(t_i|x_i,0)]^{\delta_i} S(t_i|x_i,0)} \tag{4}$$

where t_i denotes the observed death/survival time for the ith case, $f(t_i|\cdot)$ indicates the probability density of death at time t_i and $S(t_i|\cdot) = e^{-\int h(t_i|\cdot)dt}$ is the survival function for a patient who survives beyond t_i.

Sego et al. (2008) assumed the log-logistic and Weibull accelerated failure time (AFT) models for the hazard function $h(t|\cdot)$ in deriving log (LR_i) which is used to update the CUSUM. Steiner and Jones (2010) used the same parametric setting to construct the EWMA control chart. However, the distributional assumption for the hazard may be violated in some practical situations, leading to a bias in the charting statistic. Biswas and Kalbfleisch (2008) and Gandy et al. (2010) considered the Cox proportional model (2) to derive the CUSUM chart, allowing the baseline hazard to be estimated from a larger population sample.

The above methods are designed to detect changes in the average hazard. The variance, which measures the "volatility" of the surgical performance, is also important when evaluating the quality of surgery (Liu et al., 2018). For example, e.g., the adjusted risk is 10%+5% in the first case, 10%-5% in the second case, 10%+5% in the third case, and so forth. The mean value keeps 10%, but the variation suggests that the hospital may not perform stably. Therefore, the volatility of the surgical risks should be considered in the monitoring method. Instead of assuming a fixed value for γ in (2), we considered the Cox model with random effects

$$h(t|x_i) = h_0(t) e^{x_i\beta + \gamma_i} \tag{5}$$

where γ_i is a random effect for the ith patient with mean equal to zero and variance equal to σ^2. It can be interpreted as "stable" surgical performance when $\sigma = 0$, and $\sigma > 0$ indicates volatile performance, i.e., the increase of variance from zero indicates the deterioration of surgical quality. Therefore, monitoring the stability of surgical performance may be testing the hypothesis

$$H_0 : \sigma = 0 vs. H_1 : \sigma > 0 \tag{6}$$

Following Zhu and Zhang (2006), we can assume $\gamma_i = \sigma v_i$, where v_i follows an unspecified distribution F with a mean of zero and variance one. For the ith survival observations (t_i, x_i, δ_i), the conditional likelihood given random effect v_i is

$$Q_i(t_i, x_i, \delta_i | v_i) = f(t_i, x_i | v_i)^{\delta_i} S(t_i, x_i | v_i)^{1-\delta_i} = h(t_i, x_i | v_i)^{\delta_i} S(t_i, x_i | v_i) \tag{7}$$

Then the log-likelihood of the ith patient can be written as

$$l_i(\sigma) = log \int_{-\infty}^{+\infty} Q_i(t_i, x_i, \delta_i|v_i)dF(v_i)$$
$$= log \int_{-\infty}^{+\infty} [h(t_i, x_i|v_i)]^{\delta_i} S((t_i, x_i|v_i))dF(v_i) \quad (8)$$

After the integration, the likelihood is a function of σ which describes the random feature of γ_i. To perform the homogeneity test (Liang, 1987; Zhu and Zhang, 2006; Liu et al., 2018) in (6), the first-order derivative of $\sum l_i(\sigma)$ at $\sigma = 0$ can be derived as

$$S(0) = S(\sigma)|_{\sigma=0}$$
$$= 1 \Big/ 2\sum_{i=1}^{n} E_{v_i}\left[\left\{\frac{\partial log Q_i(t_i, x_i, \delta_i|v_i)}{\partial \gamma_i}\right\}^2 + \frac{\partial^2 log Q_i(t_i, x_i, \delta_i|v_i)}{\partial \gamma_i^2}\right]\Bigg|_{\sigma=0}$$
$$= 1 \Big/ 2\sum_{i=1}^{n} \{[\delta_i - \Lambda(t_i, x_i)]^2 - \Lambda(t_i, x_i)\} \quad (9)$$

where $\Lambda(t_i, x_i) = \int_0^{t_i} h_0(s)e^{x_i\beta}ds$ is the cumulative hazard for the ith case. According to the counting process theory (Andersen and Gill, 1982), the indicator $\delta_i(t)$ is a Poisson process with intensity function $\Lambda(t, x_i)$. By analogy with (Commenges and Andersen, 1995), the first part of the summation can be deemed the estimated variance. The second part is similar to the expected variance from the null hypothesis. Therefore, the null hypothesis can be rejected if the difference between the estimated and expected variances is sufficiently significant. The baseline cumulative hazard $\Lambda_0(t) = \int_0^t h_0(s)ds$ can be calculated from the historical training data (Breslow, 1972)

$$\widehat{\Lambda}_0(t) = \sum_i \frac{I(t_i \le t)\delta_i}{\sum_{j\in R_t} e^{x_j\widehat{\beta}}} \quad (10)$$

where $R_t = \{i : t_i < t\}$ denotes the risk set for those who have not died and have not been censored just prior to t. The score test statistic $S(0)$ will be identical to the results in Commenges and Andersen (1995) if the Breslow estimator (10) is included in the log-likelihood (8).

The above test statistic would be inefficient in detecting moderate and small shifts, because the likelihood function (8) ignores the information in the past observations (Zou and Tsung, 2010). In a previous study, Liu et al. (2018) used the weighted score test to monitor the surgical risk, where the fixed 30-day death risk was adjusted by logistic regression. However, using a binary outcome with a fixed cut-off time would ignore important information (Steiner and Jones, 2010). Following Zou and Tsung (2010) and Liu et al. (2018), we consider the weighted log-likelihood of the ith patient,

$$l_i(\sigma) = \lambda(1-\lambda)^{n-i} \log \int_{-\infty}^{+\infty} Q_i(t_i, x_i, \delta_i|v_i)dF(v_i) \quad (11)$$

where λ is a weighting parameter analogous to the smoothing parameter in the EWMA chart. The likelihood function indicates that the latest observations have more weights. The derivative of $\sum l_i(\sigma)$ at $\sigma = 0$ can be written as

$$S(0) = S(\sigma)|_{\sigma=0}$$
$$= \frac{1}{2}\sum_{i=1}^{n} \lambda(1-\lambda)^{n-i}\left[\left(\frac{\partial log Q_i(t_i, x_i, \delta_i|v_i)}{\partial \gamma_i}\right)^2 + \frac{\partial^2 log Q_i(t_i, x_i, \delta_i|v_i)}{\partial \gamma_i^2}\right]_{\sigma=0}$$
$$= 1/2\sum_{i=1}^{n} \lambda(1-\lambda)^{n-i}\{[\delta_i - \Lambda(t_i, x_i)]^2 - \Lambda(t_i, x_i)\}, \quad (12)$$

where $\Lambda(t_i, x_i) = \int_0^{t_i} h_0(s)e^{x_i\beta}ds$ is the cumulative hazard for the ith case, and the detailed derivation process is shown in the Appendix.

In this study, we propose the EWMA charting statistic for the nth observation (t_n, x_n, δ_n) based on $S(0)$ as

$$Z_n = \sum_{i=1}^{n} \lambda(1-\lambda)^{n-i}\{[\delta_i - \Lambda(t_i, x_i)]^2 - \Lambda(t_i, x_i)\} \quad (13)$$

which is equivalent to

$$Z_n = (1-\lambda)Z_{n-1} + \lambda Y_n, n = 1, 2, \cdots, \quad (14)$$

where $Y_n = [\delta_n - \Lambda(t_n, x_n)]^2 - \Lambda(t_n, x_n)$ and $Z_0 = 0$. We can claim that a patient's hazard is significantly different from normal (i.e., $\sigma = 0$) if the random deviation γ_i exists (i.e., $\sigma \neq 0$). Thus, the proposed risk-adjusted EWMA survival (RAES) control chart can be used to detect changes in surgical performance if the statistic Z_n is higher than an upper limit L_u or less than a lower limit L_l. Specifically, there will be a significant deterioration in surgical performance when $Z_n$is above the control limit L_u and a significant improvement when $Z_n$is below L_l. It should be noted that the signal triggered by crossing the upper limit may also indicate a change in the scale parameter because the variance can only increase from zero under the null hypothesis. When implementing the proposed control chart, the baseline hazard $\Lambda_0(t)$ and the regression parameter β can be estimated using the in-control (IC) observations in the Phase I stage (Jones-Farmer et al., 2014). Specifically, the Cox regression can be applied to IC survival outcomes and risk factors to obtain the parameter estimation of β and baseline hazard $\widehat{\Lambda}_0(t)$, in which the partial likelihood method is used to estimate the β and the Breslow's estimator $\widehat{\Lambda}_0(t)$ in (10) could be calculated at each time point. Then in phase II, the hazard $\Lambda(t_n, x_n) = \Lambda_0(t_n)e^{x_n\beta}$ and the charting statistic $Z_n$in (14) can be computed for nth observed data (t_n, x_n, δ_n). The statistic Z_n will be compared to the control limits L_u and L_l, when the nth new data is collected. The out-of-control (OC) alarm is triggered when Z_n crosses one of the control limits.

The performance of RAES charts, as proposed, is usually evaluated by the average run length (ARL) which is defined as the average number of cases required to trigger an alarm for a particular size of change. Given a pre-specified ARL, the control limits can be determined using the distribution information of the charting statistic. However, deriving the exact distribution of Z_n is intractable, even though the asymptotic distribution of Y_n can be obtained (Commenges and Andersen, 1995; Zhu and Zhang, 2006). Instead, a Monte Carlo method can be employed to determine the control limits for a given ARL. Specifically, we randomly select the IC samples based on historical data and compute the run length when an alarm is triggered. The IC ARL can be obtained by repeating the procedure 10,000 times. Then, the control limits can be determined such that the IC ARL is close to some pre-specified level by repeatedly implementing this simulation scheme with the generated data. Similar to Steiner et al. (2000), the obtained control limits are associated with the population of the risk factors. In practical use, the IC ARL may be set to some large value rather than ARL = 400, because the consequences of a false alarm may be severe.

3. Simulation study

We used a series of simulation studies to assess the performance of the proposed RAES chart with various magnitudes of shift. The underlying survival times are generated from Cox models (2) and (5) with constant baseline hazard $h_0(t) = \mu$, i.e., the baseline survival times follow an exponential distribution $\exp(\mu)$. In particular, the independent variable of the ith patient x_i is generated from the standard normal distribution and $\beta = 0.5$. Following the surgical practice for emergency patients in SOMIP, the 30-day mortality is set at 8%, and thus the parameter μ is determined so that the probability of surviving beyond 30 days is 1–8%=92%. Specifically, the random number u_i is generated from a uniform distribution in (0, 1), and the underlying true survival time for the ith patient with risk factor x_i is obtained by $t_i = \frac{u_i}{e^{x_i\beta}\log(0.92/30)}$. Further, we censor the survival time at t = 30, i.e., the observed time is t_i

if $t_i \leq 30$ and 30 if $t_i > 30$. Cox model (2) is used to generate data with different fixed shifts, and model (5) is used to simulate random shifts.

The control limits were determined using the Monte Carlo method in Section 2. In SOMIP of Hong Kong, the number of elective surgical cases in each hospital every year varies from 300 to around 1000. To control the false alarm rate for the in-control state, we set the IC ARL = 1000 in simulation studies. For comparison, we chose the RA-CUSUM chart for survival data (Biswas and Kalbfleisch, 2008; Gandy et al., 2010). The upward RA-CUSUM statistic is given by

$$Z_n + = \max(Z_{n-1} + W_n, 0) \tag{15}$$

where $Z_0 + = 0$ and W_n is determined by the log-likelihood ratio in (4). For the in-control hazard $h_0(t|x_i) = h_0(t)e^{x_i\beta}$ and out-of-control hazard $h_1(t|x_i) = \rho h_0(t|x_i)$, W_n could be rewritten as

$$W_n = \log(\rho)\delta_n - (\rho - 1)\Lambda_0(t_n)e^{x_n\beta} \tag{16}$$

where the baseline cumulative hazard at each time can be calculated by Breslow's estimator (10). The chart will signal when $Z_n +$ exceeds the upper boundary L for RA-CUSUM. For improved surgical quality, a lower-sided RA-CUSUM chart $Z_n - = \min(Z_{n-1} - W_n, 0)$ can be used. The control chart is designed to detect surgical deterioration (i.e., an increase in $Z_n +$). An improvement in surgical quality (i.e., the decrease in $Z_n -$) may still be meaningful and worth monitoring. In this way, we compare the performance of the proposed chart and the RA-CUSUM chart for both upward and downward shifts. The hazard rate model of RA-CUSUM is shown as $h(t|x_i) = \rho h_0(t|x_i)$, where the process is in-control when $\rho = 1$ and out-of-control when $\rho \neq 1$. With $\rho = e^{\gamma}$, the RA-CUSUM aims to monitor the fixed shift in (3). In the RASE chart, the random shift γ_i is monitored. Under the in-control state ($\sigma = 0$), the random shift is equivalent to the fixed shift $\gamma = 0$ because γ_i has mean zero. Therefore, the proposed RAES chart can simultaneously monitor both the scale and the location parameters.

To compare both types of charts, we used known risk-adjusted models and different λ values (0.005, 0.01, 0.05) to construct the RAES charts corresponding to $RAES^1$, $RAES^2$, $RAES^3$, respectively. Generally, small λ values are better for detecting small shifts, and large λ values are more effective for large shifts. Similar to the odds ratio setting in Liu et al.(2018), for the upward RA-CUSUM charts used to detect performance deterioration, $CUSUM^{U1}$, $CUSUM^{U2}$ and $CUSUM^{U3}$ were designed with$\rho = 1.5$, ρ=2 and$\rho = 4$, respectively. The upward CUSUM with a larger hazard ratio is expected to detect the larger shifts efficiently. For the downward RA-CUSUM charts, $CUSUM^{L1}$, $CUSUM^{L2}$, and $CUSUM^{L3}$ were designed to detect an improvement with$\rho = 0.8$, ρ=0.5 and$\rho = 0.3$, respectively. The downward CUSUM with a smaller hazard ratio is expected to detect the larger shifts efficiently. When generating data, we considered three types of shifts in the out-of-control stage: shifts in the variance of random effect—θ; a fixed shift at the intercept of the Cox model—ϵ and hazard ratios—ρ. The control limits were obtained by setting the in-control average run length to 1000 (i.e., IC ARL = 1000), and the out-of-control average run length (OC ARL) was calculated from 10,000 replications.

In practical monitoring, the hazard ratio under the alternative hypothesis and the smoothness parameter is necessary when implementing the RA-CUSUM chart and the proposed RAES chart. It could be noticed that those parameters are associated with the performance of the charts under different shift sizes. Specifically, the CUSUM type chart may perform effectively when the observed changes are identical to the hazard ratio under the alternative hypothesis. The EWMA type chart could be efficient to monitor the small changes if the small smoothing parameter is selected. However, the magnitude of the observed changes may not be precisely predicted before the chart's implementation. Thus, selecting the smoothing parameter λ in RAES and ρ in RA-CUSUM may influence the efficiency of the charts. In our study, the relative average index (RMI) value is used to evaluate the overall performance of these charts within a specific range of shift sizes, calculated as follows:

$$\text{RMI} = \frac{1}{N}\sum_{j=1}^{N}\frac{ARL_{\rho_j} - minARL_{\rho_j}}{minARL_{\rho_j}} \tag{17}$$

where N is the total number of shifts considered, ARL_{ρ_j} is the OC ARL when detecting the shift ρ_j, $minARL_{\rho_j}$ is the smallest OC ARL among all the given control charts when detecting the shift ρ_j. Thus, RMI reflects the average robustness of a control chart. According to Han and Tsung (2006), the smaller RMI indicates that the chart has an overall better performance in detecting different shift sizes.

First, we use the upward RA-CUSUM charts and the upper-sided RAES charts to detect changes in variances of random effects. From Table 1, it can be seen that the proposed RAES charts perform significantly better than the RA-CUSUM charts. The RAES charts with small smoothing parameters ($\lambda = 0.005$ and 0.01) perform better when the variances of random effect θ are small ($\theta = 0.1, 0.5, 1$). When the variance changes from 2 to 10, RAES charts with a large smoothing parameter ($\lambda = 0.05$) detect the shifts more efficiently. Regarding the RMI values, the RAES chart with $\lambda = 0.005$ outperforms others over different magnitudes of shift, suggesting that it can be used to detect practical "instability" of surgical risks. In addition, the upward RA-CUSUM with$\rho = 2$ has the smallest RMI value of the three RA-CUSUM charts, indicating it could be the choice if the RA-CUSUM chart is expected to monitor variance in a practical implementation.

We also consider the shifts with different ratios of hazards. For example, ρ=2 means that the OC hazard is twice the IC hazard. $\rho > 1$ is equivalent to a deterioration in surgical performance, which should be monitored by the upward RA-CUSUM chart. Conversely, the case where $\rho < 1$ should be monitored by the downward RA-CUSUM chart. The results for upward control charts and downward controls are included in Tables 2 and 3, respectively. The ARL results show that small shifts can be detected more quickly by RAES charts with small λ values. As ρ increases, the RA-CUSUM charts perform better when the shifts match those they are designed for. For example, $CUSUM^{U3}$ whose parameterρ = 4 performs best when the shift ratio is exactly equal to 4, where ARL = 20.28. It can be observed that RAES charts with small λ ($RAES^1$) can detect small shifts (ρ from 0.5 to 2) more efficiently than RA-CUSUM charts. When shifts become large (ρ = 0.1,0.3,4,8), RA-CUSUM charts show better performance than RAES charts. By comparing the RMI results, we conclude that a RAES chart with $\lambda = 0.005$ is most robust for different changes in hazard ratios. Thus the $RAES^1$ chart can be used to monitor practical surgical performance when the real changes are unknown.

Finally, we compare the performance of these charts when the fixed shifts occur at the intercept of the Cox model, i.e., ϵ from $h(t|x_i) = h_0(t)e^{x\beta+\epsilon}$, at different values. The results are shown in Tables 4 and 5, where Table 4 compares performance when monitoring the deterioration of surgical risk, and Table 5 records the results of detecting improvements in surgical performance. The RAES charts with small λshow the lowest ARL values when the shifts, ϵ, are small ($-1 \leq \epsilon < 1$), indicating that $RAES^1$ can detect small changes in hazard more efficiently than RA-CUSUM charts. On the other hand, the RA-CUSUM designed with$\rho = 4$ performs better when shifts are large ($|\epsilon| > 1$), consistent with the previous simulation results. Regarding the RMI, the RAES chart

Table 1
ARL and RMI performance of the upward CUSUM and the RAES charts with different θ values.

	$\theta = 0.1$	$\theta = 0.5$	$\theta = 1$	$\theta = 2$	$\theta = 5$	$\theta = 10$	RMI
$CUSUM^{U1}$	984	544	170.1	43.83	19.20	15.19	0.2232
$CUSUM^{U2}$	976	578	179.1	40.68	16.62	13.11	0.1680
$CUSUM^{U3}$	972	642	228.8	42.43	14.51	11.98	0.2146
$RAES^1$	966	466	139.2	39.33	17.29	13.60	0.0882
$RAES^2$	968	503	147.3	39.30	16.63	13.12	0.0960
$RAES^3$	973	606	193.2	38.37	14.01	10.71	0.1159

Table 2
ARL and RMI performance of the upward CUSUM and the RAES charts with different ρ values.

	$\rho = 1.1$	$\rho = 1.3$	$\rho = 1.5$	$\rho = 2$	$\rho = 4$	$\rho = 8$	RMI
$CUSUM^{U1}$	589	264.8	154.7	72.46	24.67	12.83	0.2098
$CUSUM^{U2}$	620	288.0	164.6	71.36	21.75	11.13	0.1935
$CUSUM^{U3}$	674	363.6	219.0	88.84	20.28	9.38	0.3398
$RAES^{1}$	504	216.9	131.3	65.46	23.79	12.61	0.0862
$RAES^{2}$	532	236.4	138.2	67.31	23.24	12.20	0.1121
$RAES^{3}$	641	320.8	188.7	78.46	20.50	9.88	0.2418

Table 3
ARL and RMI performance of the downward CUSUM and the RAES charts with different ρ values.

	$\rho =$ 0.9	$\rho =$ 0.8	$\rho =$ 0.7	$\rho =$ 0.5	$\rho =$ 0.3	$\rho =$ 0.1	RMI
$CUSUM^{L1}$	566	354.4	245.8	135.8	98.36	73.90	0.2093
$CUSUM^{L2}$	614	382.3	252.7	132.1	81.22	57.34	0.1511
$CUSUM^{L3}$	634	416.4	276.9	142.8	78.74	51.76	0.1873
$RAES^{1}$	511	303.1	209.3	120.4	81.48	61.72	0.0379
$RAES^{2}$	548	333.3	221.4	120.9	79.03	58.02	0.0598
$RAES^{3}$	668	449.3	307.9	155.8	87.85	54.69	0.2878

Table 4
ARL and RMI performance of the upward CUSUM and the RAES charts with different ϵ values.

	$\epsilon =$ 0.1	$\epsilon = 0.3$	$\epsilon = 0.5$	$\epsilon = 0.8$	$\epsilon = 1$	$\epsilon = 2$	RMI
$CUSUM^{U1}$	577	227.3	115.9	58.56	41.16	13.74	0.1776
$CUSUM^{U2}$	602	248.1	121.8	56.06	38.03	11.79	0.1604
$CUSUM^{U3}$	660	314.7	159.6	66.46	40.94	10.03	0.3168
$RAES^{1}$	490	186.2	102.6	53.96	39.24	13.30	0.0596
$RAES^{2}$	520	204.0	107.4	54.38	38.46	12.85	0.0840
$RAES^{3}$	623	275.6	136.4	59.40	38.72	10.59	0.2093

Table 5
ARL and RMI performance of the downward CUSUM and the RAES charts with different ϵ values.

	$\epsilon =$ -0.1	$\epsilon =$ -0.3	$\epsilon =$ -0.5	$\epsilon =$ -0.8	$\epsilon =$ -1.0	$\epsilon =$ -2.0	RMI
$CUSUM^{L1}$	584	282.2	183.6	129.2	109.6	77.6	0.2167
$CUSUM^{L2}$	625	297.6	181.3	115.2	93.7	60.6	0.1355
$CUSUM^{L3}$	650	324.2	196.0	117.9	93.0	55.2	0.1643
$RAES^{1}$	521	242.5	153.9	108.1	91.2	64.9	0.0321
$RAES^{2}$	567	258.6	161.9	107.5	90.2	61.0	0.0520
$RAES^{3}$	678	356.0	220.5	134.0	105.2	59.3	0.2815

with$\lambda = 0.005$ shows the greatest robustness. The RA-CUSUM designed with$\rho = 2$ and 0.5 has the best performance of the RA-CUSUM charts, suggesting that the $RAES^{1}$, upward $CUSUM^{L2}$, and downward $CUSUM^{L2}$ should be the choices when implementing practical monitoring.

4. Application to the real data

This section applies the proposed RAES and the RA-CUSUM charts to a real data series from the Surgical Outcome Monitoring and Improvement Program (SOMIP) in Hong Kong. Since the actual size of the shift in the real data is unknown, we chose the RAES chart with parameter$\lambda = 0.005$, the upward CUSUM chart with$\rho = 2$, and the downward CUSUM chart with$\rho = 0.5$ to detect both deterioration and improvement in surgical performance, based on the RMI results of simulation studies. The data include the surgical outcomes from several hospitals between July 2009 and June 2012. For example, the data set from hospital A has the results of 1378 emergency surgical cases, including 894 operations from July 2009 to June 2011 and 484 operations in the 2011–2012 period.

First, we monitored hospital A's quality of surgery in 2011–2012 based on its previous average performance during 2009–2011. In particular, we used the data during 2009–2011 to train the risk-adjusted model, and the RAES chart was applied to monitor the surgical outcomes during 2011–2012. From the clinical practice in SOMIP, several risk factors, such as, Conscious of Neurological Status, Impaired Sensorium of Neurological Status, Presence of Uncorrected Bleeding, Use of Oral or Parenteral Steroid, Preoperative Chemotherapy for Malignancy, etc., were selected in the Cox regression model to adjust the survival outcomes. The obtained risk-adjusted hazard model was

$$\begin{aligned} h(t) = h_0(t)exp(&\text{Conscious of Neurological Status} * (-0.399) \\ &+ \text{Impaired Sensorium of Neurological Status} * 0.255 \\ &+ \text{Presence of Uncorrected Bleeding} * 1.05 \\ &+ \text{Use of Oral or Parenteral Steroid} * 0.215 \\ &+ \text{Preoperative Chemotherapy for Malignancy} * 0.692 + \ldots) \end{aligned}$$

The baseline cumulative hazard $\Lambda_0(t) = \int_0^t h_0(s)ds$ can also be estimated by using the Breslow estimator in (10). Then the Monte Carlo method in Section 2 was employed to determine the control limits. Specifically, a bootstrap procedure was used to randomly draw the IC samples in 2009–2011, and the control charts were applied to those samples to calculate the run length. We repeated this step 10,000 times to compute the ARL. Finally, the control limits were selected to fulfill the pre-specified IC ARL. For hospital A, the IC ARL is set as 600, which is slightly higher than the yearly average of surgical cases in hospital A, to obtain the control limit. Thus, the number of false alarms is less than one for the monitoring period if no change occurred. In addition, the annual number of emergency surgery from all hospitals is around 8000, and thus the IC-ARL = 8000 was also used to determine the control limits.

The RAES and CUSUM charts are shown in Fig. 1. It can be observed that both RAES and CUSUM charts describe an improvement in hospital A in the early stages (before the 100th case) compared to its two-year average. The early improvement detected by the RAES chart triggers the alarm faster than the downward RA-CUSUM chart, indicating that the improvement may be subtle if the alarm signal is true. During the late stage of the monitoring period (after the 400th case), another improvement is detected by the RAES chart but not captured by the RA-CUSUM. This may be due to practical improvements in the surgical performance often being slightly over a one-year period.

In addition, the same risk adjustment was applied to the cases of all hospitals in 2009–2011, and the RAES chart and RA-CUSUM charts were used to monitor whether hospital A's performance exceeded the average of all hospitals in the period 2011–2012. Then we determined the upper and lower control limits using data from the previous two years and IC ARL = 8,000 because the yearly average of all hospitals is around 8000. The results are shown in Fig. 2. A downward trend was also identified in the early stage (before the 100 cases) by both the RAES and RA-CUSUM charts. However, this early improvement does not trigger an alarm, which may be because the magnitude is small compared to the average performance of all hospitals. After the 420th case, an improvement is detected by the RAES chart but not by the RA-CUSUM chart. This may be because the proposed RAES chart can detect small changes more efficiently than the RA-CUSUM chart if the alarm signal at the late stage is accurate.

5. Conclusions

In this paper, we propose a control chart for continuous survival time after surgery, aiming to monitor the average surgical risk and the existence of its variance. By applying the Cox model, the baseline hazard function can be estimated nonparametrically and avoid the Sego et al. (2009) distributional assumption. The random effects added in the Cox model are used to describe shifts in the hazard function, and then a

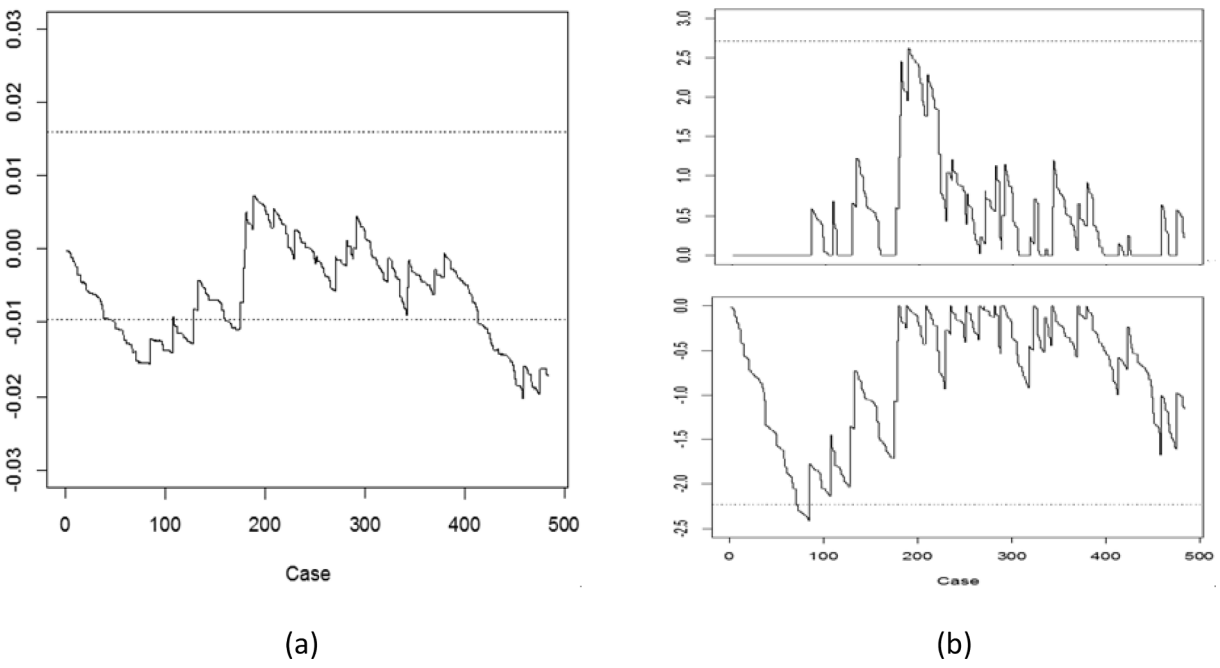

Fig. 1. RAES (a) and CUSUM (b) charts monitor the deviation from hospital A's average performance.

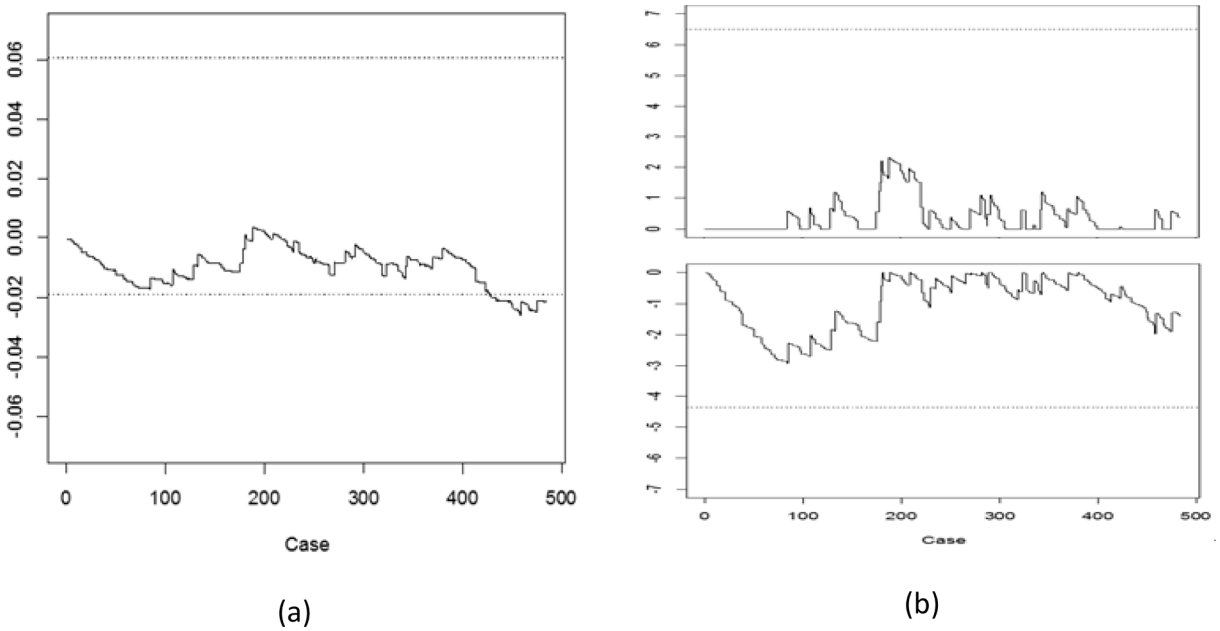

Fig. 2. RAES (a) and CUSUM (b) charts for hospital A.

score-type statistic is obtained to test whether the variance of random effects is significant or not by using the homogeneity test (Liang, 1987; Zhu and Zhang, 2006; Liu et al., 2018). Because the variance under the null hypothesis ($\sigma = 0$) indicates that random effects in the hazard are not significant, the proposed RAES chart applying the EWMA procedure can be used to monitor shifts of both the average level and the variance parameter sequentially. For surgical cases, the proposed method could be employed to detect the stability of the surgical performance and changes in the average surgical risk. In the experimental part, we carried out a series of simulation studies and real data applications and took the RA-CUSUM chart in Biswas and Kalbfleisch (2008) for comparison. To assess the performance of the above two charts with different magnitude of shifts in simulations, we performed experiments changing the variance of random effects, the fixed shifts at the intercept of the Cox model, and the hazard ratios. The results show that the proposed RAES chart can detect shifts in the variance more efficiently than the RA-CUSUM chart, confirming the ability of the RAES chart to monitor the scale parameter. The RAES chart performs better for the location parameter than the RA-CUSUM chart in detecting small shifts, and the RA-CUSUM is superior for large shifts. By comparing the RMI values, the RAES chart with $\lambda = 0.005$ is most robust within the range of shifts, suggesting that this chart should monitor changes in surgical performance because the shifts that occur are usually unknown before implementation. Comparing to the RA-CUSUM chart, which aims to detect the changes in average risk, the proposed RAES chart can monitor the existence of variance in addition to the mean values. In a practical

surgical risk monitoring project, the hospital authority could be interested in the methods that can identify the clinical root causes when the proposed chart triggered an alarm.

It should be noted that the post-diagnostic method after alarm should further combine with the clinical experience to help discover the problem. When applied to real surgical data in Hong Kong, the proposed RAES appears more sensitive to the shifts during the monitoring period. This may be because practical changes in surgical performance are often subtle over one year. Further, the chart statistics can adopt the daily updating scheme in Gan et al.(2020) to avoid the delay caused by patient orders. In summary, we have extended surgical performance monitoring control charts for survival outcomes by adapting the score test for random effects. As a result, the detection of changes in the average hazard and its "stability" can be achieved. Both the simulation and practical studies verified the excellent performance of the proposed RAES chart. In this study, the monitoring method was derived from the null hypothesis in which the scale parameter is zero. Because it is also meaningful to know whether the volatility is less than a certain threshold in practice, the approach based on the null hypothesis with nonzero scale parameter is needed in the future study.

Acknowledgment

The authors thank editor, associate editor and the two anonymous referees for their helpful comments that have resulted in significant improvements to the article. The work is supported by grants from National Natural Science Foundation of China (No. 71872146, 71931006, 12075162), the RGC General Research Fund (16216119, 16201718), Sichuan Science and Technology Department (No. 2020YJ0357), and VC&VR Key Lab of Sichuan Province.

Data Availability

The datasets are obtained from the surgical outcomes monitoring and improvement project (SOMIP) in the Hospital Authority (HA) of Hong Kong. The datasets generated and analyzed during the current study are not publicly available due to that we have signed a confidentiality agreement with the Hospital Authority of Hong Kong. However, the data used for research purposes are available from the SOMIP committee in the Hospital Authority of Hong Kong upon reasonable request. For those who are interested, please contact the author or refer to the website of HA https://www3.ha.org.hk/data/DCL/Index/.

CRediT authorship contribution statement

Xin Lai: Conceptualization, Methodology. **Xiao Li:** Conceptualization, Methodology. **Liu Liu:** Methodology. **Fugee Tsung:** Methodology. **Paul B.S. Lai:** . **Jiayin Wang:** . **Xuanping Zhang:** . **Xiaoyan Zhu:** Data curation. **Jiaqi Liu:** Data curation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix:

The demonstration of the equation is as follows:

Let $\theta = \sigma^2$, then we can rewrite the $l_i(\sigma)$as

$$l_i(\theta) = \lambda(1-\lambda)^{n-i} log \int_{-\infty}^{\infty} Q_i(t_i, x_i, \delta_i|v_i)\mathrm{dF}(v_i). \tag{A.1}$$

So we can find that

$$S(\theta) = \sum_{i=1}^{n} \lambda(1-\lambda)^{n-i}\frac{\partial l_i(\theta)}{\partial\theta} = \sum_{i=1}^{n} \lambda(1-\lambda)^{n-i}\frac{\frac{\partial \int_{-\infty}^{\infty} Q_i(\theta)\mathrm{d}F(v_i)}{\partial\theta}}{\int_{-\infty}^{\infty} Q_i(\theta)\mathrm{d}F(v_i)}, \tag{A.2}$$

where we can simply write $Q_i(t_i,x_i,\delta_i|v_i)$as $Q_i(\theta)$. For all of the v_i obey the same distribution, we can simplify the question to consider only a specific v. if we use the Taylor expansion at a certain point θ_0 to get the expansion form $Q(\theta)$, we can show that

$$\begin{aligned} Q(\theta) &= Q(\theta_0) + \frac{v\theta_0^{-\frac{1}{2}}}{2}\frac{\partial Q(\theta_0)}{\partial\gamma}(\theta-\theta_0) \\ &+ \left[\frac{\partial^2 Q(\theta_0)}{\partial\gamma^2}v^2\theta_0^{-1} - \frac{\partial Q(\theta_0)}{\partial\gamma}v\theta_0^{-\frac{3}{2}}\right]\frac{(\theta-\theta_0)^2}{8}. \end{aligned} \tag{A.3}$$

Therefore,

$$\begin{aligned} \int_{-\infty}^{\infty} Q(\theta)\mathrm{d}F(v) &= \int_{-\infty}^{\infty} \left[\begin{array}{c} Q(\theta_0) + \frac{v\theta_0^{-\frac{1}{2}}}{2}\frac{\partial Q(\theta_0)}{\partial\gamma}(\theta-\theta_0) + \\ \left[\frac{\partial^2 Q(\theta_0)}{\partial\gamma^2}v^2\theta_0^{-1} - \frac{\partial Q(\theta_0)}{\partial\gamma}v\theta_0^{-\frac{3}{2}}\right]\frac{(\theta-\theta_0)^2}{8} \end{array}\right]\mathrm{d}F(v) \\ &= Q(\theta_0) + \frac{\partial^2 Q(\theta_0)}{\partial\gamma^2}\frac{(\theta-\theta_0)^2}{8\theta_0}. \end{aligned} \tag{A.4}$$

Since the mean of v is 0 and the variance of v is 1, we have $\int_{-\infty}^{\infty} v\mathrm{d}F(v) = 0$, $\int_{-\infty}^{\infty} v_2\mathrm{d}F(v) = 1$. That is why we got the equation (A.4). We can easily deduce that

$$\frac{\partial \int_{-\infty}^{\infty} Q\mathrm{d}F(v)}{\partial\theta} = \frac{\partial^2 Q(\theta_0)}{\partial\gamma^2}\frac{\theta-\theta_0}{4\theta_0}. \tag{A.5}$$

Then we will figure out $S(\theta)|_{\theta=0}$. In this condition, we know that $0 < \theta_0 < \theta$ and θ is really close to 0, so we can use L'Hospital's rule to calculate the equation (A.5). If we fix $\theta = k\theta_0$, we have

$$\lim_{\substack{\theta\to 0\\ \theta=k\theta_0}} \frac{\partial \int_{-\infty}^{\infty} Q(\theta)\,\mathrm{d}F(v)}{\partial\theta} = \lim_{\substack{\theta\to 0\\ \theta=k\theta_0}} \frac{\partial^2 Q(\theta_0)}{\partial\gamma^2}\frac{\theta-\theta_0}{4\theta_0} = \frac{k-1}{4}\frac{\partial^2 Q(\theta)}{\partial\gamma^2}. \tag{A.6}$$

If we set k = 3, there will be

$$\lim_{\substack{\theta\to 0\\ \theta=3\theta_0}} \frac{\frac{\partial \int_{-\infty}^{\infty} Q(\theta)\,\mathrm{d}F(v)}{\partial\theta}}{\int_{-\infty}^{\infty} Q(\theta)\,\mathrm{d}F(v)} = \lim_{\substack{\theta\to 0\\ \theta=3\theta_0}} \frac{\frac{\partial^2 Q(\theta_0)}{\partial\gamma^2}\frac{\theta-\theta_0}{4\theta_0}}{Q(\theta_0)} = \frac{\frac{\partial^2 Q(\theta)}{\partial\gamma^2}}{2Q(\theta)}|_{\theta=0}. \tag{A.7}$$

Therefore,

$$S(\theta)|_{\theta=0} = \frac{1}{2}\sum_{i=1}^{n}\lambda(1-\lambda)^{n-i}\frac{\frac{\partial^2 Q_i(t_i,x_i,\delta_i|v_i)}{\partial\gamma^2}}{Q_i(t_i,x_i,\delta_i|v_i)}|_{\theta=0}. \tag{A.8}$$

We can quickly know the truth that

$$\begin{aligned}
&\left(\frac{\partial log Q_i(t_i,x_i,\delta_i|v_i)}{\partial\gamma_i}\right)^2 + \frac{\partial^2 log Q_i(t_i,x_i,\delta_i|v_i)}{\partial\gamma_i^2}\\
&= \left(\frac{\frac{\partial Q_i(t_i,x_i,\delta_i|v_i)}{\partial\gamma_i}}{Q_i(t_i,x_i,\delta_i|v_i)}\right)^2 + \frac{\frac{\partial^2 Q_i(t_i,x_i,\delta_i|v_i)}{\partial\gamma_i^2}}{Q_i(t_i,x_i,\delta_i|v_i)} - \left(\frac{\frac{\partial Q_i(t_i,x_i,\delta_i|v_i)}{\partial\gamma_i}}{Q_i(t_i,x_i,\delta_i|v_i)}\right)^2\\
&= \frac{\frac{\partial^2 Q_i(t_i,x_i,\delta_i|v_i)}{\partial\gamma_i^2}}{Q_i(t_i,x_i,\delta_i|v_i)}.
\end{aligned} \tag{A.9}$$

Therefore, $S(\theta)$ can be rewritten as

$$\begin{aligned}
&S(\theta)|_{\theta=0}\\
&= \frac{1}{2}\sum_{i=1}^{n}\lambda(1-\lambda)^{n-i}\left[\left(\frac{\partial log Q_i(t_i,x_i,\delta_i|v_i)}{\partial\gamma_i}\right)^2 + \frac{\partial^2 log Q_i(t_i,x_i,\delta_i|v_i)}{\partial\gamma_i^2}\right]_{\theta=0}\\
&= 1/2\sum_{i=1}^{n}\lambda(1-\lambda)^{n-i}\left\{[\delta_i - \Lambda(t_i,x_i)]^2 - \Lambda(t_i,x_i)\right\},
\end{aligned} \tag{A.10}$$

where $\Lambda(t_i,x_i) = \int_0^{t_i} h_0(s)e^{x_i\beta}ds$. For $\theta = \sigma^2$, if the $\theta\to 0$, then $\sigma\to 0$, we can also find that

$$\begin{aligned}
&S(\sigma)|_{\sigma=0}\\
&= \frac{1}{2}\sum_{i=1}^{n}\lambda(1-\lambda)^{n-i}\left[\left(\frac{\partial log Q_i(t_i,x_i,\delta_i|v_i)}{\partial\gamma_i}\right)^2 + \frac{\partial^2 log Q_i(t_i,x_i,\delta_i|v_i)}{\partial\gamma_i^2}\right]_{\sigma=0}\\
&= 1/2\sum_{i=1}^{n}\lambda(1-\lambda)^{n-i}\left\{[\delta_i - \Lambda(t_i,x_i)]^2 - \Lambda(t_i,x_i)\right\},
\end{aligned} \tag{A.11}$$

References:

Andersen, P. K., & Gill, R. D. (1982). Cox's regression model for counting processes: A large sample study. *Annals of Statistics, 10*, 1100–1120.

Biswas, P., & Kalbfleisch, J. D. (2008). A risk-adjusted CUSUM in continuous time based on the Cox model. *Statistics in Medicine, 27*(17), 3382–3406.

Breslow, N. E. (1972). Discussion of the paper by D. R. *Cox. J R Statist Soc B, 34*, 216–217.

Castagliola, P., & Tsung, F. (2005). Autocorrelated SPC for non-normal situations. *Quality and Reliability Engineering International, 21*(2), 131–161.

Chen, Z., Lu, S., & Lam, S. (2007). A hybrid system for SPC concurrent pattern recognition. *Advanced Engineering Informatics, 21*(3), 303–310.

Commenges, D., & Andersen, P. K. (1995). Score test of homogeneity for survival data. *Lifetime Data Analysis, 1*(2), 145–156.

Cook, D. A., Coory, M., & Webster, R. A. (2011). Exponentially weighted moving average charts to compare observed and expected values for monitoring risk-adjusted hospital indicators. *BMJ Quality and Safety, 20*(6), 469–474.

Cook, D. A., Duke, G., Hart, G. K., Pilcher, D., & Mullany, D. (2008). Review of the Application of Risk-Adjusted Charts to Analyze Mortality Outcomes in Critical Care. *Critical Care Resuscitation, 10*(3), 239–251.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statisitcal Society: Series B, 34*, 187–202.

Gan, F. F., Yuen, J. S., & Knoth, S. (2020). Quicker Detection Risk-Adjusted Cumulative Sum Charting Procedures. *Statistics in medicine, 39*(7), 875–889.

Gandy, A., Kvaloy, J. T., Bottle, A., & Zhou, F. (2010). Risk-Adjusted Monitoring of Time to Event. *Biometrika, 97*(2), 375–388.

Grigg, O. AJ. (2019). The STRAND Chart: A survival time control chart. *Statistics in Medicine, 38*(9), 1651–1661.

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

Grigg, O., & Spiegelhalter, D. J. (2007). Simple risk-adjusted exponentially weighted moving average. *Journal of the American Statistical Association, 102*, 140–152.

Han, D., & Tsung, F. (2006). A reference-free cuscore chart for dynamic mean change detection and a unified framework for charting performance comparison. *Journal of the American Statistical Association, 101*(473), 368–386.

In, H., Palis, B.E., Merkow, R.P., Posner, M.C., Ferguson, M.K., Winchester, D.P., Pezzi, C. M. (2016) Doubling of 30-Day Mortality by 90 Days after esophagectomy: a critical measure of outcomes for quality improvement. Ann Surg 263: 286-291.

Jiang, W., & Tsui, K. L. (2008). A theoretical framework and efficiency study of multivariate statistical process control charts. *IIE Transactions, 40*(7), 650–663.

Jones-Farmer, L. A., Woodall, W. H., Steiner, S. H., & Champ, C. W. (2014). An overview of Phase I analysis for process improvement and monitoring. *Journal of Quality Technology, 46*(3), 265–280.

Keefe, M. J., Loda, J. B., Elhabashy, A. E., & Woodall, W. H. (2017). Improved implementation of the risk-adjusted Bernoulli CUSUM chart to monitor surgical outcome quality. *International Journal for Quality in Health Care, 29*(3), 343–348.

LIANG, KUNG.-YEE. (1987). A Locally Most Powerful Test for Homogeneity with Many Strata. *Biometrika, 74*(2), 259–264.

Liu, L., Lai, X., Zhang, J., & Tsung, F. (2018). Online profile monitoring for surgical outcomes using a weighted score test. *Journal of Quality Technology, 50*(1), 88–97.

Lovegrove, J., Valencia, O., Treasure, T., Sherlaw-Johnson, C., & Gallivan, S. (1997). Monitoring the Results of Cardiac Surgery by Variable Life-Adjusted Display. *Lancet, 350*(9085), 1128–1130.

Mant, J. (2001). Process versus outcome indicators in the assessment of quality of health care. *International Journal for Quality in Health Care, 13*(6), 475–480.

Merkow, R. P., Bilimoria, K. Y., & Ko, C. Y. (2013). Surgical quality measurement: An evolving science. *JAMA Surg, 148*(7), 586. https://doi.org/10.1001/jamasurg.2013.128

Mise, Y., Vauthey, J.N., Zimmitti, G., Parker, N.H., Conrad, C., Aloia, T.A., Lee, J.E., Fleming, J.B., Katz, M.H. (2015) Ninety-day postoperative mortality is a legitimate measure of hepatopancreatobiliary surgical quality. Ann Surg 262: 1071-1078.

Paynabar, K., Jin, J.(., & Yeh, A. B. (2012). Phase I risk-adjusted control charts for monitoring surgical performance by considering categorical covariates. *Journal of Quality Technology, 44*(1), 39–53.

Shang, Y., Tsung, F., & Zou, C. (2013). Statistical process control for multistage processes with binary outputs. *IIE transactions, 45*(9), 1008–1023.

Shen, X., Zou, C., Jiang, W., & Tsung, F. (2013). Monitoring Poisson count data with probability control limits when sample sizes are time varying. *Naval Research Logistics (NRL), 60*(8), 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), 61–65.

Sego, L. H., Reynolds, M. R., & Woodall, W. H. (2009). Risk-adjusted monitoring of survival times. *Statistics in Medicine, 28*(9), 1386–1401.

Shu, L., & Tsung, F. (2003). On multistage statistical process control. *Journal of the Chinese Institute of Industrial Engineers, 20*(1), 1–8.

Steiner, S. H., Cook, R. J., Farewell, V. T., & Treasure, T. (2000). Monitoring surgical performance using risk-adjusted cumulative sum charts. *Biostatistics, 1*, 441–452.

Steiner, S. H., & Jones, M. (2010). Risk-adjusted survival time monitoring with an updating exponentially weighted moving average (EWMA) control Chart. *Statistics in Medicine, 29*(4), 444–454.

Steiner, S.H. (2014) 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, 245-264.

Talsma, A. K., Linqsma, H. F., Steyerberg, E. W., et al. (2014). The 30-day versus in-hospital and 90-day mortality after esophagectomy as indicators for quality of care. *Ann Surg., 260*, 267–273.

Treasure, T., Taylor, K., & Black, N. (1997). *Independent Review of Adult Cardiac Surgery-United Bristol*. Bristol: Health Care Trust.

Waldie, P. (1998) Crisis in the Cardiac Unit. The Globe and Mail, Canada's National Newspaper, Oct. 27, Section A: 3, column 1.

Walters, D. M., McMurry, T. L., Isbell, J. M., Stukenborg, G. J., & Kozower, B. D. (2014). Understanding mortality as a quality indicator after esophagectomy. *Ann Thorac Surg., 98*(2), 506–512.

Woodall, W. H., Fogel, S. L., & Steiner, S. H. (2015). The monitoring and improvement of surgical outcome quality. *Journal of Quality Technology, 47*(4), 383–399.

Yue, J., Lai, X., Liu, L., & Lai, P. B. S. (2017). A new VLAD-based control chart for detecting surgical outcomes. *Statistics in Medicine, 36*(28), 4540–4547.

Yuen W.C. (2013). Applying Variable Life Adjusted Display in Monitoring Surgical Outcomes. https://www3.ha.org.hk/haconvention/hac2013/proceedings/downloads/MC1.1.pdf.

Zhu, H., & Zhang, H. (2006). Generalized Score Test of Homogeneity for Mixed Effects Models. *Annals of Statistics, 34*, 1545–1569.

Zou, C., & Tsung, F. (2010). Likelihood ratio-based distribution-free EWMA control charts. *Journal of Quality Technology, 42*(2), 174–196.