

Risk-adjusted survival time monitoring with an updating exponentially weighted moving average (EWMA) control chart

Stefan H. Steiner^{a*†} and Mark Jones^b

Monitoring medical outcomes is desirable to help quickly detect performance changes. Previous applications have focused mostly on binary outcomes, such as 30-day mortality after surgery. However, in many applications the survival time data are routinely collected. In this paper, we propose an updating exponentially weighted moving average (EWMA) control chart to monitor risk-adjusted survival times. The updating EWMA (uEWMA) operates in a continuous time; hence, the scores for each patient always reflect the most up-to-date information. The uEWMA can be implemented based on a variety of survival-time models and can be set up to provide an ongoing estimate of a clinically interpretable average patient score. The efficiency of the uEWMA is shown to compare favorably with the competing methods. Copyright © 2009 John Wiley & Sons, Ltd.

Keywords: control chart; monitoring; risk adjustment; survival time

1. Introduction

There is a clear need for monitoring the outcomes of medical processes such as surgery. Timely signals of poor performance can be used to improve the process and avoid future unnecessary deaths or other adverse outcomes. With a sound statistical approach, we can balance the desire for quick detection of problems with the need to avoid frequent false alarms.

Process monitoring has a long history in the industry starting with Shewhart [1]. Some early applications of control charts in medicine, such as De Leval *et al.* [2] and Steiner *et al.* [3], who considered the problem of monitoring paediatric cardiac surgery outcomes, did not employ risk adjustment. This was reasonable in their context as none of the possible covariates collected, such as patient characteristics, procedural characteristics and surgical team fatigue, were found to have a significant effect on the failure rate. However, in most medical applications, due to the heterogeneity of patients, risk adjustment is necessary. Otherwise, a short series of adverse outcomes for a cluster of high-risk patients could falsely signal a systematic problem.

A variety of risk-adjusted monitoring procedures have been proposed. Lovegrove *et al.* [4, 5] and Poloniecki *et al.* [6] suggest simple monitoring schemes based on a plot of the difference between the cumulative predicted and the observed deaths. Steiner *et al.* [7] proposed a risk-adjusted cumulative sum (CUSUM) control chart based on a binary outcome. The risk-adjusted CUSUM uses optimal likelihood ratio weights and has been shown to be more efficient than methods based on predicted minus expected deaths, but may be difficult for clinicians to interpret. For further discussion see the review papers by Grigg and Farewell [8] and Woodall [9].

Most previous medical applications of statistical process monitoring have focused on binary outcomes, such as 30-day mortality. However, in many medical applications, survival time data, such as time to death or major complication, time to infection, etc. are routinely collected. In fact, many previous applications, such as Steiner *et al.* [7], use examples where the collected survival time data are discretized into 30-day mortality. Such discretization results in a loss of information and a delay in accounting for outcomes. In many applications it makes sense to monitor the survival time data directly.

Two recent papers address monitoring survival time (or time to failure) data with risk adjustment. Sego *et al.* [10] propose a risk-adjusted survival time (RAST) CUSUM based on log-likelihood ratio scores. They illustrate the approach with log-logistic and Weibull accelerated failure time (AFT) models but other models are also possible. To ensure that the patients remain in

^aDepartment of Statistics and Actuarial Sciences, University of Waterloo, Waterloo, Ont., Canada N2L 3G1

^bCentre for Healthcare Related Infection Surveillance and Prevention (CHRISP), Queensland Health, Herston, Queensland 4006, Australia

*Correspondence to: Stefan H. Steiner, Department of Statistics and Actuarial Sciences, University of Waterloo, Waterloo, Ont., Canada N2L 3G1.

†E-mail: shsteine@uwaterloo.ca

surgery order, the chart is updated only after each patient reaches 30 days post surgery. At that time, any patient who died before 30 days receives the appropriate log-likelihood score based on their time of death, whereas surviving patients receive the appropriate score for an observation censored at 30 days. This approach delays the entry of early failures into the CUSUM, but unfortunately if we allow early failures to enter the RAST CUSUM immediately while successes are not identified until the 30-day mark, we have the potential to mix up patient ordering before and after a process change.

Biswas and Kalbfleisch [11] propose a risk-adjusted CUSUM in continuous time. They illustrate the approach with kidney transplantation data and use a Cox regression model to make the necessary risk adjustments, although other survival-time models are possible. The approach involves the continuous inclusion of likelihood ratio scores based on Bernoulli outcomes. While the Biswas CUSUM is considerably more complex than the Sego CUSUM, the approach is appealing as new information is incorporated into the chart immediately. However, the CUSUM statistic is not easily interpretable and may be hard to promote to clinicians.

For monitoring survival time data, we propose an updating exponentially weighted moving average (uEWMA) control chart. EWMA control charts are similar to CUSUM charts in the sense that both accumulate information across multiple time periods to look for process changes. Such sequential control charts are good at detecting small persistent changes in the process. CUSUM and EWMA control charts are generally believed to have similar efficiency [12]. The main advantage of an EWMA is that it provides an ongoing local estimate of the average score. In this way, it is easier for clinical staff to interpret and understand. Another minor advantage is the inherent two-sided nature of an EWMA. Tabular CUSUM schemes, such as suggested by Steiner *et al.* [7] and Sego *et al.* [10], on the other hand, are designed to detect one-sided process changes, although they can be made two-sided through the simultaneous use of two CUSUMs (with different scores).

In Section 2 we briefly review the standard EWMA chart and define the continuous time uEWMA. We discuss the approach in the context of monitoring surgical survival times, but the method is also applicable in other contexts with survival time data. In addition, to simplify the language, we equate an adverse outcome with death, though it could be something else, for example a major complication or an infection. We describe how the uEWMA can be updated on an ongoing basis to reflect the latest information. In Section 3 we suggest three types of possible patient scores, including optimal log-likelihood ratio scores, observed minus expected deaths scores and scores based on a summary of the survival-time distribution. In Section 4 we apply the proposed uEWMA to a cardiac surgery example and illustrate how to perform the necessary calculations and produce the uEWMA control chart. Note that in this paper we show a retrospective application of the chart for illustration. Ultimately, the chart would be used prospectively in real time. Section 5 addresses the design of the uEWMA control chart and shows how the design parameters affect the theoretical efficiency of the chart and Section 6 compares the uEWMA with the Sego RAST CUSUM [10] and the Biswas CUSUM [11]. Finally, in Section 7 we draw the conclusions and provide a discussion of some additional issues.

2. Risk-adjusted uEWMA

The standard EWMA control chart [13] is given by

$$\begin{aligned} E_i &= \gamma s_i + (1 - \gamma)E_{i-1} \\ &= \gamma s_i + \gamma(1 - \gamma)s_{i-1} + \gamma(1 - \gamma)^2 s_{i-2} + \gamma(1 - \gamma)^3 s_{i-3} + \dots \end{aligned}$$

where γ is a smoothing constant, $0 < \gamma \leq 1$. In our context, s_i is a score assigned to patient i and E_0 equals some suitable starting value that we set equal to the (estimated) average patient score before any process change. We shall refer to the functions of γ as the patient weights, e.g. the weight for patient $i-2$ is $\gamma(1-\gamma)^2$. To create the control chart we plot E_i versus time (actually patient number), and the chart signals if $E_i > h_U$ or $E_i < h_L$, where h_U and h_L are pre-specified constants, called the upper and lower control limits, respectively. We anticipate that when the EWMA is used to monitor the process performance prospectively, a chart signal would trigger an investigation to try to determine the cause of the suspected change in the performance. We shall set the control limits to balance the desire to quickly detect real process changes while avoiding frequent false alarms where the process has not changed, but where the chart signals due to chance alone.

In the standard EWMA, as E_i is a weighted average, it is an estimate of the average score that gives more weight to patients who have had surgery more recently. Thus, EWMA signals, where $E_i > h_U$, suggest an increase in the average patient score, whereas $E_i < h_L$ suggests that the average score has decreased. In what follows we set the patient scores such that increases in the average score correspond to the worsening surgical performance. This matches clinicians' expectations as they are used to monitoring failure or mortality rates.

EWMA control charts have been used extensively for monitoring industrial processes and Morton *et al.* [14], Spliid [15] and Grigg and Spiegelhalter [16] provide examples of its use in a medical context. In addition, in time series analysis the EWMA has been employed as a smoother and as a one-step ahead forecast for a special case of an autoregressive integrated moving average model [17].

We propose adapting the standard EWMA to a continuous time scale by allowing the patient scores to be functions of time. This allows the latest information about all patients to be incorporated in the EWMA at all times. To accomplish this we change the notation to focus on time (in addition to patient number). We define the uEWMA as

$$E_t = \gamma s_{it} + \gamma(1 - \gamma)s_{i-1,t} + \gamma(1 - \gamma)^2 s_{i-2,t} + \gamma(1 - \gamma)^3 s_{i-3,t} + \dots \quad (1)$$

where s_{it} is the score for patient i (where the index i gives the order of surgery) at time t . Note that in (1), given time t , the value of i is determined. Patient i corresponds to the most recent patient to have had surgery, that is at time t , i patients have had surgery. It is important that (1) preserves the patients in surgery order because we believe that patient outcomes depend on the surgical quality. Our goal is to quickly detect changes in the surgical performance which we believe could be sudden but sustained. Our focus here is mainly on detecting deterioration of the surgical performance. As a result, in this paper, rather than using a standard lower control limit (h_L), we will instead implement a reflecting lower control limit. That is the EWMA will not be allowed to go below h_L , but hitting h_L will not reset the EWMA. We shall set the reflecting barrier so that when the in-control model parameters have been well estimated its effect is very small, but the reflecting barrier will provide protection against chart inertia [18, 19]. That is the reflecting barrier will limit the amount of credit the chart can build up due to a sequence of good performances. Too much built up credit just before the deterioration in performance would delay the signal. If during implementation the EWMA hits the reflecting barrier repeatedly it is a sign that the in-control parameter estimates are no longer or never were good estimates of the true process performance. In this case we recommend re-estimating the in-control parameters.

As we describe in the next section, we can employ a variety of patient scores. We illustrate with scores based on the log-likelihood ratio, observed minus expected deaths and a local (based on the results for only patient i) estimate of some summary of the survival-time distribution for a baseline or typical patient. With these latter scores we follow Grigg and Spiegelhalter [16] who suggested something similar when monitoring a binary outcome with an EWMA. The survival model summary scores can be selected to represent a variety of survival-time distribution summaries. We shall illustrate using the 30-day mortality rate (as estimated from the survival model), but there are many other possibilities.

In application, the uEWMA (1) will be updated whenever there is a failure (i.e. death) and also at some regular time interval, say each week. In this way we will always quickly detect when there is evidence that the surgical performance has deteriorated and have a (nearly) up-to-date estimate of the average patient score. At each updating time, we recalculate E_t using the latest scores for each patient (and the appropriate corresponding weights that depend on each patient's surgery order). As time passes, more patients undergo surgery, and the past patient's EWMA weights will gradually decrease exponentially as given in (1). The new value of E_t would then be plotted. Note that once a value of E_t has been plotted it is never updated. The idea is that E_t reflects an estimate of the average score at time t , and thus it is not updated as more information becomes available at later times. Thus, the updating of patient scores affects only the future values of E_t .

As the number of patients increases the weights for early patients become very small. Thus, in implementation, patients from long ago can be ignored. To be conservative in our simulation studies described later, we ignore patients when their weight is less than $\frac{1}{1000}$ th the magnitude of the weight for the most recent patient, that is when $\gamma(1-\gamma)^i < \gamma/1000$. When $\gamma=0.01$ this works out to $i > 688$ and we need only keep track of the weights and scores for the most recent 688 patients. In many examples this can be further reduced as the patient scores can no longer change after they have reached the maximum follow-up time, say 30 days. Thus, for all patients whose surgery was over 30 days ago we only need to keep track of the corresponding EWMA summary statistic E (for these patients).

The efficiency of a sequential control chart, like an EWMA, is most commonly described by the run length distribution. Run length is defined as the time (or number of patients) until the chart signals. Typically, the run length distribution is summarized using the average run length (ARL). However, run length distributions are bounded by zero and usually have long right tails. As a result, the ARL is difficult to estimate with simulation and may not summarize the efficiency well. In this paper we will assess the efficiency instead with the median run length which is easier to simulate and gives arguably a better summary.

Generally, application of a control chart such as an EWMA chart is divided into two phases. In phase I we set up the control chart by selecting the design parameters, e.g. γ , h_U and h_L . This is typically done using the phase I data, which we believe represent an in-control process where the performance is constant. The control chart is then retrospectively applied to the phase I data to check the in-control assumption. In phase II we apply the designed control chart on an ongoing basis to new data. In phase II, a signal would lead to the investigation of the possible reasons for the changed (better or worse) performance. Timely intervention could save lives or prevent unnecessary adverse events. We design the uEWMA to look for changes in the mean (or median) survival time that manifest themselves through changes in the scale/location parameter of the survival-time distribution. We assume that all other parameters, such as the shape parameter and the effect of any covariates, are constant.

For monitoring survival time data where, as in the applications/examples discussed in this paper, the observed mortality rate is low, small values of the EWMA smoothing constant (γ), say around 0.01, work best. Small values of γ imply a lot of smoothing as each observation has a small weight. The proposed smoothing constant is smaller than the typically recommended values for EWMA's used in the industry [12]. In most industrial applications, EWMA's are based on subgroups and scores are calculated as a summary of a number of observations. We address the choice of smoothing constant further in Section 5.

3. Determining the patient scores

To derive the scores we need to first define some notation for the observed data for each patient. At time t , for patient i (that has had surgery), we observe $(x_{it}, \delta_{it}, \mathbf{u}_i)$, where x_{it} is the minimum of the current time since time zero, the time to death and the follow-up time (or time at occurrence of a competing risk) each minus the time of surgery,

$$\delta_{it} = \begin{cases} 1 & \text{if patient } i \text{ dies by time } t \\ 0 & \text{otherwise} \end{cases}$$

and \mathbf{u}_i is a vector of covariate values. Note that the covariate values for each patient are determined at the time of surgery and are not updated as time passes. In many applications, such as the cardiac surgery example presented in Section 4, the follow-up time is fixed and is the same for all patients. The follow-up time is defined as the time after which any adverse outcome can no longer be reasonably attributable to the surgery or simply the maximum time for which we can reliably collect the outcome data. For instance, if the outcome is time to infection after surgery, after a patient has been discharged from the hospital it is much harder to determine whether a patient acquired an infection or not.

To be more precise, for patient i we denote t as the current time, a_i as the time of surgery, c_i as the time of a competing risk (or follow-up time) and d_i as the time of a death. Then $x_{it} = \min(t, c_i, d_i) - a_i$. Note that c_i and d_i represent realizations of random variables only the smaller of which is observed. In Section 4 our example will illustrate the methodology with a fixed follow-up time, i.e. $c_i = a_i + c$, where c is some fixed constant, say 30 days. However, in the general case c_i may be the time of a competing risk such as some other adverse outcome unrelated to the surgery or a variable follow-up time.

Thus, for patient i (that has had surgery) there are three possibilities for (x_{it}, δ_{it}) :

- (i) Death: $(x_{it}, 1)$, where $x_{it} = d_i - a_i$ is the time between surgery and death.
- (ii) Success: $(x_{it}, 0)$, where $x_{it} = c_i - a_i$ is the time between surgery and the follow-up time (or some competing risk).
- (iii) At risk: $(x_{it}, 0)$, where $x_{it} = t - a_i$ is the time between surgery and the current time.

The patient scores, s_{it} , are based on $(x_{it}, \delta_{it}, \mathbf{u}_i)$; hence, as x_{it} and possibly δ_{it} change for case (iii) as time passes, so will (some of) the scores. A patient in case (iii) can become case (i) or (ii) or remain in case (iii) with a larger x_{it} . Note that once a patient is in case (i) or (ii) x_{it} and δ_{it} (and thus the patient score) stay the same.

The patient scores also depend on the selected survival-time distribution. In this context we need a model that allows for risk adjustment, such as an AFT model [20] or a Cox regression. There are many possible survival-time distributions, including Weibull, log normal, etc. To illustrate we present the results based on the AFT log-logistic distribution. The popular AFT models assume that the survivor function for a patient with covariate vector \mathbf{u} at time x is the same as the baseline survivor function at time $x \exp(\beta^T \mathbf{u})$, where β is a vector of covariate coefficients, i.e. the covariate effect is log-linear in time. The baseline survivor function is determined by the chosen survival-time distribution. The AFT log-logistic model for a patient with covariates \mathbf{u} has probability density function

$$f(x) = \frac{\alpha}{\lambda} (x \exp(\beta^T \mathbf{u}) / \lambda)^{\alpha-1} [1 + (x \exp(\beta^T \mathbf{u}) / \lambda)^\alpha]^{-2} \quad (2)$$

where α and λ are the shape and scale parameters, respectively. The corresponding survivor function is

$$\Pr(X > x) = S(x|U = u) = [1 + (x \exp(\beta^T \mathbf{u}) / \lambda)^\alpha]^{-1} \quad (3)$$

We use the phase I data to estimate the AFT log-logistic parameters. In what follows, we denote the phase I (in-control) parameter estimates as α_0 and λ_0 and assume that they are estimated without error. Of course, there is some error that depends on the size and quality of the phase I data. See Sego *et al.* [10] for more on the effect of estimation error on the efficiency of the RAST CUSUM, and Jones *et al.* [21] and Jensen *et al.* [22] for more on this issue in general, when there is no risk adjustment, on the efficiency of the standard EWMA and other control charts.

For the log-logistic AFT model, a patient with a covariate vector \mathbf{u} has a median and mean survival time given by

$$\begin{aligned} \text{Median} &= \lambda / \exp(\beta^T \mathbf{u}) \\ \text{Mean} &= \begin{cases} \lambda \alpha \sin(\pi / \alpha) / [\pi \exp(\beta^T \mathbf{u})] & \text{if } \alpha > 1 \\ \text{undefined} & \text{otherwise} \end{cases} \end{aligned} \quad (4)$$

As mentioned earlier, we use the uEWMA to look for changes in the mean or median which are due to changes in the scale parameter λ while assuming that the shape parameter α is fixed at its in-control value α_0 , similar to Sego *et al.* [10].

Next, we look at three classes of patient scores, namely log-likelihood ratio scores, observed minus expected deaths scores and survival model summary scores. The log-likelihood ratio scores are preferred, because as shown in Section 6, they provide the most power to detect process changes. The observed minus expected and survival model summary scores have the advantage that the resulting uEWMA will, on an ongoing basis, provide a local estimate of a clinically interpretable average score. This should appeal to practitioners who prefer monitoring approaches that are transparent and easy to understand.

3.1. Log-likelihood ratio scores

The log-likelihood ratio scores are based on the methodology of Steiner *et al.* [7] and Sego *et al.* [10]. Moustakides [23] showed that in the non-risk-adjusted case a CUSUM with log-likelihood ratio scores is optimal such that they will give the smallest ARL (when the shift takes place at the worst possible time and when the chart is furthest away from a signal) at the specified process shift while maintaining the in-control ARL.

We set up the uEWMA to detect changes in the log-logistic scale parameter as shifts in λ correspond directly to shifts in the average or median survival time, as seen in (4). Consider the hypothesis test:

$$H_0: \lambda = \lambda_0 \quad \text{versus} \quad H_1: \lambda = \rho_1 \lambda_0 \quad (5)$$

where we assume that the log-logistic shape parameter (α) is unchanged. Then, following Sego *et al.* [10], the log-likelihood scores for the i th patient at time t for the log-logistic AFT model are given by (6) where the observed patient values ($x_{it}, \delta_{it}, \mathbf{u}_i$) are defined in Section 3, λ_0 and α_0 are the in-control log-logistic scale and shape parameters, respectively, and β is a vector of parameters for the covariates

$$s_{it} = -\alpha \delta_{it} \log(\rho_1) + 2^{\delta_{it}} \left\{ \log \left[1 + \left(\frac{x_{it} \exp(\beta^T \mathbf{u}_i)}{\lambda_0} \right)^{\alpha_0} \right] - \log \left[1 + \left(\frac{x_{it} \exp(\beta^T \mathbf{u}_i)}{\rho_1 \lambda_0} \right)^{\alpha_0} \right] \right\} \quad (6)$$

3.2. Observed minus expected deaths scores

The use of the observed minus expected deaths scores is popular when monitoring binary outcomes [4–6]. Here we apply the same methodology to the monitoring of survival time data. For each patient the expected death rate is given by the cumulative distribution function. For instance, for the AFT log-logistic model, given the covariates \mathbf{u}_i , the chance of death by time x_{it} is one minus the survivor function (3). Thus, the observed minus expected deaths scores for patient i based on the AFT log-logistic model are given by (7)

$$s_{it} = \delta_{it} - (1 - [1 + (x_{it} \exp(\beta^T \mathbf{u}_i) / \lambda_0)^{\alpha_0}]^{-1}) \quad (7)$$

Note that in (7) we use the observed minus expected scores rather than the other way around as is popular in some applications [4–6]. With observed minus expected scores, increases suggest a deterioration in the surgical performance.

3.3. Survival model summary scores

There are two steps required to determine the survival model summary score s_{it} for patient i at time t .

1. Translate the outcome for patient i , i.e. $(x_{it}, \delta_{it}, \mathbf{u}_i)$, into something equivalent for the typical (or benchmark) patient, denoted (y_{it}, δ_{it}) .
2. Translate (y_{it}, δ_{it}) from Step 1 into a local (based on the outcome for patient i) estimate for the chosen summary of the lifetime distribution (in terms of the baseline/typical patient).

To accomplish the risk adjustment in Step 1, we need to define a typical or baseline patient. We denote the covariate values for the typical/baseline patient with the vector \mathbf{u}_b . We choose \mathbf{u}_b in consultation with the clinicians who will use the control chart. For instance the typical patient may be the ‘average’ or most common type of patient. Then in Step 1, we find y_{it} by solving $S(x_{it} | \mathbf{u}_i) = S(y_{it} | \mathbf{u}_b)$, where \mathbf{u}_i gives the covariate values for patient i . In other words, we match results between patient i and the typical patient using the tail probabilities of the survivor (or equivalently the cumulative distribution) function. For instance, using the log-logistic AFT survivor function given by (3) we obtain

$$y_{it} = x_{it} \exp(\beta^T \mathbf{u}_i) / \exp(\beta^T \mathbf{u}_b) \quad (8)$$

In the second step to determine the survival model summary scores, we need to do three things:

- 2(a) Translate y_{it} given by (8) into an estimate of the central location, e.g. mean or median, of the selected failure time distribution.
- 2(b) Derive a method of moments estimate of the location parameter of the selected survival-time distribution, denoted λ_{it} based on the estimate of the central location from 2a.
- 2(c) Find the score s_{it} by translating λ_{it} into an estimate of the selected survival distribution summary.

For step 2(a), if $\delta_{it} = 1$, i.e. the patient died, we estimate the central location as y_{it} . For censored observations, i.e. $\delta_{it} = 0$, on the other hand, we estimate the central location using conditioning. We determine either the conditional expected value as suggested by Steiner [24] for monitoring the censored observations or the conditional median. For the AFT log-logistic model the conditional mean is

$$E(X|X > y_{it}, \mathbf{u}_b, \alpha_0, \lambda_0) = \int_{y_{it}}^{\infty} xf(x) dx$$

where $f(x)$ is given in (2). The conditional mean is only defined if $\alpha > 1$ and there is no closed-form solution, but it can be found using numerical integration. The conditional median for the AFT log-logistic distribution is given by

$$\text{median}(X|X > y_{it}, \mathbf{u}_b, \alpha_0, \lambda_0) = \frac{\lambda_0}{\exp(\beta^T \mathbf{u}_b)} \left(1 + 2 \left[\frac{y_{it} \exp(\beta^T \mathbf{u}_b)}{\lambda_0} \right]^{\alpha_0} \right)^{1/\alpha_0}$$

In Step 2(b) we determine a method of moments estimate of the location parameter. We do this by assuming the shape parameter is unchanged from the in-control estimate. To illustrate, we use the median, as in the example in the next section where $\alpha_0 < 1$ the log-logistic mean is undefined. Using the median, for the log-logistic AFT model with $\alpha = \alpha_0$, our local estimate of λ , for

patient i at time t , is

$$\lambda_{it} = \begin{cases} y_{it} \exp(\beta^T \mathbf{u}_b) & \text{if } \delta_{it} = 1 \\ \text{median}(X|X > y_{it}, \mathbf{u}_b, \alpha_0, \lambda_0) \exp(\beta^T \mathbf{u}_b) & \text{if } \delta_{it} = 0 \end{cases} \quad (9)$$

In Step 2(c), based on α_0 and λ_{it} , as given by (7), we derive a local estimate of the selected distributional summary (for the baseline patient). There are many possible choices of distributional summary including failure rate at some time, average or median failure time, expected survival time for the first 10 per cent of failed patients, cumulative hazard rate at some time after surgery, etc. To illustrate, (10) gives scores for the 30-day mortality rate for the baseline patient (estimated from the survival-time distribution) from the AFT log-logistic model

$$s_{it} = 1 - \left[1 + \left(\frac{30 \exp(\beta^T \mathbf{u}_b)}{\lambda_{it}} \right)^{\alpha_0} \right]^{-1}$$

$$= \begin{cases} 1 - \left[1 + \left(\frac{30 \exp(\beta^T \mathbf{u}_b)}{x_{it} \exp(\beta^T \mathbf{u}_i)} \right)^{\alpha_0} \right]^{-1} & \text{if } \delta_{it} = 1 \\ 1 - \left[1 + \frac{30 \exp(\beta^T \mathbf{u}_b)}{\lambda_0} \left[1 + 2 \left(\frac{x_{it} \exp(\beta^T \mathbf{u}_i)}{\lambda_0} \right)^{\alpha_0} \right]^{-1/\alpha_0} \right] & \text{if } \delta_{it} = 0 \end{cases} \quad (10)$$

4. Cardiac surgery example

To illustrate we apply the uEWMA to the cardiac surgery example also considered in Steiner *et al.* [7] and Segó *et al.* [10]. The data set consists of 6994 operations, from a single surgical center over the seven-year period, 1992–1998. Information about each patient included date, surgeon, type of procedure and the pre-operative variables that comprise the Parsonnet score [25]. These include age, gender, hypertension, diabetic status, renal function and left ventricular mass. In the data, 461 deaths occurred within 30 days of surgery, giving an overall 30-day mortality rate of 6.6 per cent. We fix the follow-up time as 30 days, as after 30 days we believe that the outcome may no longer be attributable to the surgery quality.

As in Steiner *et al.* [7], we select the first two years of data (corresponding to 1992 and 1993) as the phase I (in-control) period. Fitting the log-logistic AFT model as done by Segó *et al.* [10] gives the maximum likelihood estimates $(\alpha_0, \lambda_0, \beta) = (0.529, 30606, 0.145)$. Note that in this example the only covariate is the Parsonnet value. As we have $\beta = 0.145$, the median time to death is higher for larger Parsonnet values, as we would expect.

For illustration, in Table I, we show the patient scores for a variety of example patients. For the survival model summary scores, we selected the typical/baseline patient as the median risk phase I patient. Thus, $u_b = 7$ since this is the median Parsonnet value in the phase I data. In addition, as in this example $\alpha_0 < 1$ the log-logistic mean is not defined and we use the conditional median rather than the mean in Step 2(a). Note from Table I that the scores change little for low-risk patients who survive longer but a lot for high-risk patients who survive longer.

Using the methodology shown in the next section, for the log-likelihood ratio scores we select a control limit h_U of 0.022, a smoothing constant (γ) of 0.01 and an initial EWMA value (E_0) of -0.015 . In addition, to protect against chart inertia we select a reflecting barrier h_L of -0.04 . Figure 1 gives the resulting uEWMA for both the phase I and the phase II data with $\rho_1 = 0.2697$, as in Segó *et al.* [10], which corresponds to setting the alternative hypothesis to a doubling in the odds of 30-day mortality. In the figure we multiply the control limit h_U by the factor $1 - (1 - \gamma)^{2i}$ to take into account the effect of the initial EWMA value in the variability of the EWMA statistic [26]. With the large number of patients the control limit quickly asymptotes to 0.022.

$(x_{it}, \delta_{it}, u_i)$	s_{it} (6)	s_{it} (7)	y_{it}	λ_{it}	s_{it} (10)	Comment
(10, 1, 1)	0.66	0.991	4.2	11.56	0.829	Low risk, died early
(10, 0, 1)	-0.015	-0.009	4.2	30 329	0.070	Low risk, short survival
(30, 0, 1)	-0.027	-0.016	12.6	30 375	0.070	Low risk, long survival
(10, 1, 10)	0.63	0.982	15.4	42.63	0.708	Medium risk, died early
(10, 0, 10)	-0.03	-0.018	15.4	30 391	0.070	Medium risk, short survival
(30, 0, 10)	-0.05	-0.031	46.4	30 562	0.070	Medium risk, long survival
(10, 1, 50)	0.02	0.720	5103	14 081	0.101	High risk, died early
(10, 0, 50)	-0.34	-0.280	5103	60 965	0.049	High risk, short survival
(30, 0, 50)	-0.43	-0.411	15 309	131 518	0.033	High risk, long survival

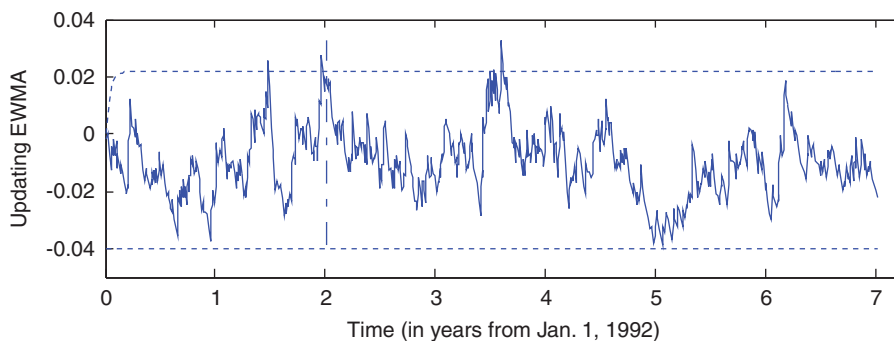


Figure 1. Updating EWMA applied to SGH data with log-likelihood ratio scores given by (6). Upper dashed (horizontal) line gives the control limit, lower dashed line gives the reflecting barrier, dash-dot (vertical) line divides Phase I from Phase II.

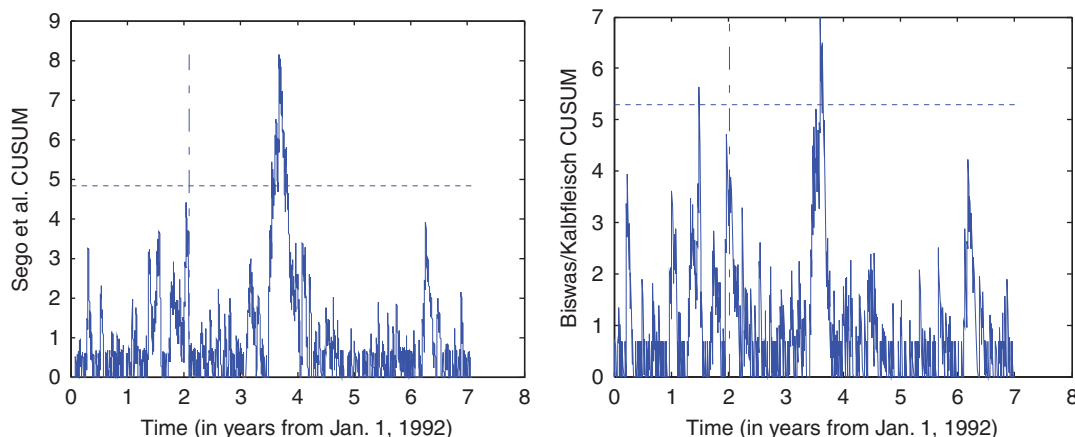


Figure 2. Figure 2: Sego RAST and Biswas CUSUM for the cardiac surgery example. Dashed line gives the control limit, dash-dot line divides Phase I from Phase II.

For comparison, Figure 2 shows the Sego RAST and Biswas CUSUMs applied to the same cardiac surgery data and assuming the same log-logistic AFT model. For both the Sego RAST and Biswas CUSUMs, we also selected the alternative hypothesis (5) as a doubling in the odds of 30-day mortality. For all the control charts in Figures 1 and 2 the control limits were set to give a median in-control run length of roughly 3000 days. This closely matches the results with the control limit choice of 4.8327 in Sego *et al.* [10]. From Figures 1 and 2 we see that the Sego RAST CUSUM has a very pronounced signal in phase II, whereas the Biswas CUSUM and uEWMA show similar patterns and signal only very briefly in both phases I and II. All charts signal in phase II around the middle of the fourth year (i.e. in 1995). Had any of these charts been used prospectively for the phase II data the signaled process changes would have initiated an investigation into what, if anything, had changed. However, looking back retrospectively we see that if there were any process changes they were not sustained. We conclude that the signals were most likely false alarms. Owing to the large number of surgeries represented by this data having some false alarms is not unexpected.

The differences in the nature of the patterns seen in Figures 1 and 2 are due to the differences in how the patient results are entered into the CUSUM or EWMA calculations. With the Biswas and uEWMA, deaths are accounted for on the actual day of death, whereas with the Sego approach deaths always enter the CUSUM 30 days after surgery. The larger signal in the Sego RAST CUSUM in 1995 is due to a cluster of deaths of patients whose surgeries were very close together. In the uEWMA and Biswas CUSUM, on the other hand, the effect of these deaths is muted because the actual days of death are spread out, with information from surviving or at-risk patients appearing in the gaps.

5. Designing an updating EWMA chart

Once we have chosen the appropriate survival-time distribution, important covariates and the scores, to design the uEWMA chart we need to choose the smoothing constant γ and the control limits h_U and h_L .

We evaluate the choices based on the run length distribution determined through Monte Carlo simulation. Usually the run length distribution of an EWMA can be determined numerically using a Markov chain-based approximation as discussed in Lucas and Saccucci [12] and Steiner [26]. However, with the uEWMA all scores may be updated at each step and thus the Markovian assumption does not hold. The simulations are computationally intensive; hence, we summarize the run length distribution with the easier-to-calculate median run length rather than the ARL.

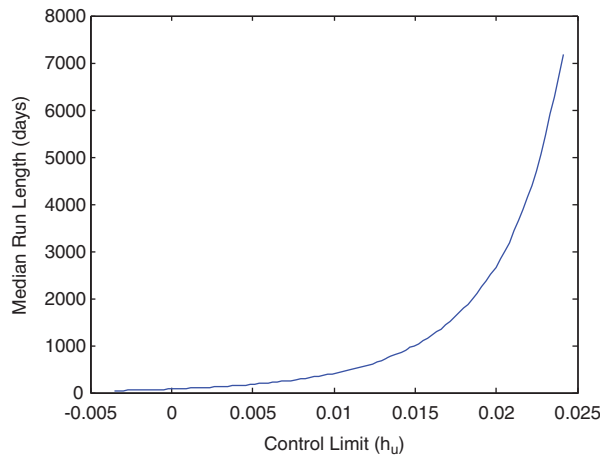


Figure 3. In-control simulated median run length as a function of control limit (h_U). Using log-likelihood ratio weights given by (6) with $\rho_1 = 0.2697$.

To conduct the simulation we need to make some assumptions about the patients. In particular, we need a model for the

- Survival-time distribution.
- Distribution of time to a competing risk (or follow-up time).
- Arrival pattern and distribution of the covariates (i.e. patient mix).

We base our simulation on the cardiac surgery example described in Section 4. We assume that the survival-time distribution is given by the AFT log-logistic model fit to the phase I data. The follow-up time is 30 days for all patients and we assume an exponential time between patients (i.e. patients arrive according to a homogenous Poisson process) with an average of 2.7 per day as this is the rate for the years 1992–1993 (i.e. in phase I). Finally, as in Sego *et al.* [10], we assume that the Parsonnet covariates follow an exponential distribution with mean 8.9. This fits the observed pattern in the phase I data fairly well.

In situations where a competing risk can occur or if the follow-up time is variable we could model the times using the empirical rates given for all the patients in phase I who did not die. The time to a competing risk could be modeled using the same or different covariates rather than the survival-time distribution.

In all cases, when running the simulation we first generate a number of patients from the in-control distribution. The number of in-control patients depends on the smoothing constant γ and is chosen so that the initial E_0 value has a weight that is less than $\frac{1}{1000}$ th of the weight for the next patient. With $\gamma = 0.01$ this translates to 688 in-control patients. These patients are used to provide a random initial condition for the EWMA and the CUSUMs at the time of the simulated process change. As a consequence, the simulated results are steady-state run lengths rather than zero state run lengths. Steady-state run lengths are more realistic as we cannot predict when the surgical performance may change. After the initial group of in-control patients, we simulate a second group of enough patients to determine the median run length. With these data we generate the corresponding EWMA. This process is repeated 20 000 times to approximate the median run length. Only signals resulting from the second group of patients are used to estimate the run length distribution.

Figure 3 shows the median run length as a function of the upper control limit (h_U) when the process remains in-control, that is we have $\lambda = 30606$ (and $\alpha = 0.529$). Here we used the log-likelihood ratio scores with $\rho_1 = 0.2697$, as in Sego *et al.* [10], which corresponds to a doubling in the odds of 30-day mortality given the assumed patient mix. Figure 3 was used to set a reasonable control limit. For instance with a control limit (h_U) of 0.022, we estimate that the median run length is about 3000 (days) or 8100 surgical procedures. That is if nothing has changed we expect a false alarm from the EWMA to occur before 8.2 years half the time. We can select a different control limit if we want a higher or lower median run length. From the results used to create Figure 3 we note that the natural logarithm of the median run length is roughly a linear function of the control limit. We could also explore other summaries of the run length distribution if desired. MATLAB (www.mathworks.com) code to simulate the median run length is available upon request from the first author. The choice of the control limit must be made carefully. Setting it to give very frequent false alarms will render the monitoring method useless as practitioners will learn to ignore it.

To determine the results in this section and the next we use a lower reflecting barrier h_L of -0.04 . The results in Figure 3 would change for larger values of h_L in that we would need a larger upper control limit h_U to give the same in-control median run length. However, the efficiency results in Figures 4–5 and Table II would be virtually unchanged for larger h_L values.

To simulate the out-of-control behavior of the uEWMA suppose the log-logistic scale parameter shifts to $\lambda_1 = \lambda_0/q$, where λ_0 is the in-control value, i.e. when $q = 2$ the median survival time has been halved. We can quantify how the median run length changes as we increase q by repeating the simulation described earlier when the second group of patients has a survival-time distribution with scale parameter λ_1 . To place these changes into perspective, from the log-logistic survivor function (3), a change in the odds of mortality from O_0 to fO_0 corresponds to $q = f^{1/\alpha_0}$, where α_0 is the in-control shape parameter estimate. For example, with $\alpha_0 = 0.529$ a doubling in the odds of 30-day mortality, i.e. $f = 2$, corresponds to $q = 3.7$ (Note $q = 1/\rho_1$ where ρ_1 is defined in (5)).

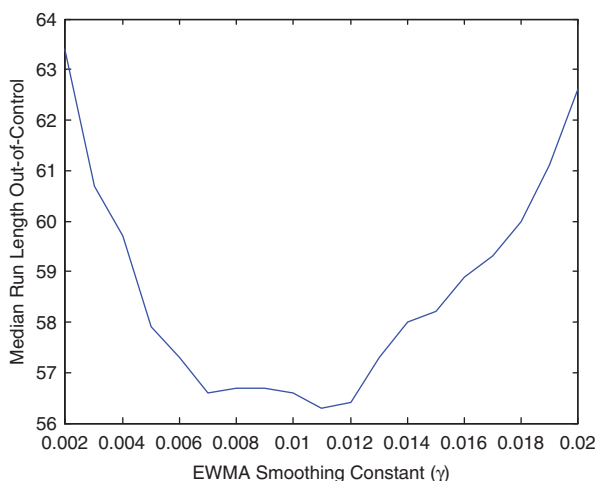


Figure 4. Effect of changing the EWMA smoothing constant (γ). In-control median run length=1000, out-of-control mean shifted to $q=3.7$.

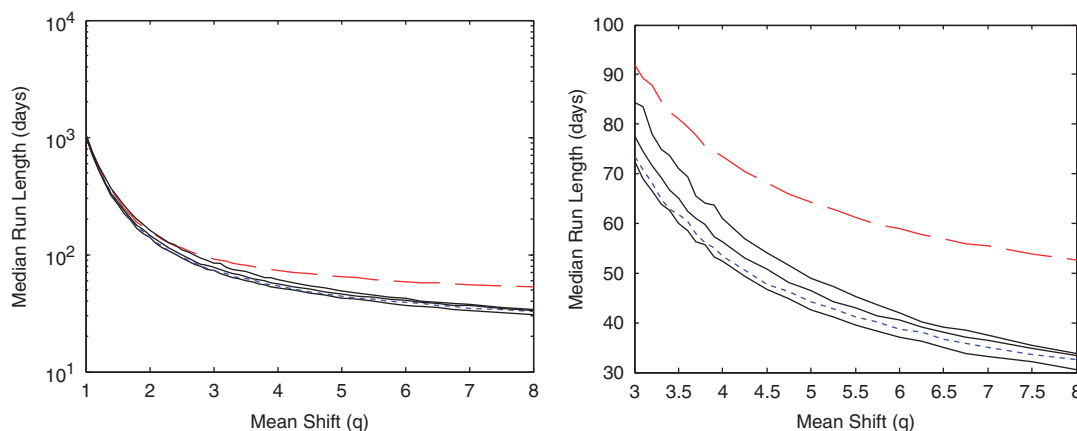


Figure 5. Median run lengths as a function of a mean shift uEWMA with scores (6), (7) and (10) (solid lines ordered from bottom to top) Biswas CUSUM (dotted line) and Sego CUSUM (dashed line).

Table II. Comparison of monitoring methods.		
Method	Median run length (days) at $q=3.7$	Approx. number of extra adverse events
uEWMA with log-likelihood scores	56.4	10.0
Biswas CUSUM	58.0	10.3
uEWMA with observed-expected scores	60.4	10.7
uEWMA with 30-day mortality scores	62.2	11.1
Sego CUSUM	77.3	13.8

We have illustrated the use of the uEWMA with a smoothing constant $\gamma=0.01$. Other choices are possible. Figure 4 compares the various choices. In all cases the control limit was selected so that the in-control median run length was approximately 1000 days. We expect the relative efficiency given the various choices for γ not to depend on the choice of the in-control median run length. The plot gives the corresponding out-of-control median run length when $q=3.7$. As shorter out-of-control median run lengths are better, a value near 0.01 seems optimal. This is a similar value as found by Cook [27] in the context of monitoring binary outcomes with an EWMA chart. Note that this result is somewhat dependent on the failure rate in the example. In the presented cardiac surgery example, few patients (6.6 per cent) died before reaching the maximum follow-up time of 30 days. In examples with higher failure rates we expect that the best value of the EWMA smoothing constant would be something higher than 0.01.

6. Comparison of efficiency

In this section we compare the efficiency of the uEWMA with the alternatives suggested by Sego *et al.* [10] and Biswas and Kalbfleisch [11] using simulated data modeled on the cardiac surgery example discussed in Section 4. To make the methods comparable, in all the cases we used the same AFT log-logistic survival-time distribution and make the same assumptions about the patient arrival rate and patient mix. In addition, for all approaches the likelihood ratio scores corresponded to a doubling in the odds of 30-day mortality. Note that with the Sego RAST CUSUM a new patient is added only 30 days after surgery. This inherent delay affects the efficiency.

In the comparison of the three methods, we set the control limits in each case so that the in-control median run length was 1000 days. This means that we expect a false alarm within roughly three years 50 per cent of the time. The selected control limits are for the uEWMA with log-likelihood ratio scores $h_U=0.0150$, the uEWMA with observed minus expected deaths scores $h_U=0.0554$, the uEWMA with 30-day mortality scores $h_U=0.0908$, the Biswas CUSUM $h=4.1525$ and the Sego RAST CUSUM $h=3.9636$. In this comparison, for all the uEWMA's we do not use a lower reflecting control limit. This choice has no effect on the results, but makes the simulations run faster.

Figure 5 shows the median run length as a function of the mean shift, q , for the uEWMA with log-likelihood ratio scores (6), observed minus expected deaths scores (7) and the 30-day mortality scores (10) for the cardiac surgery example when the in-control median run length is 1000 days. We selected 1000 days rather than 3000 days (as in Figure 1) to reduce the computational burden of the simulations. The relative efficiency of the various charts should not depend on the choice of the in-control median run length. We see that, for this example, the uEWMA with log-likelihood ratio scores is the most efficient; its median run length for all process shifts is the smallest. The efficiency of the Biswas CUSUM is slightly worse especially for larger process shifts. The other two uEWMA charts, i.e. with observed minus expected deaths scores and 30-day mortality rate scores, have very similar but a little worse efficiency than the Biswas CUSUM, whereas the inherent delay results in the Sego RAST CUSUM having the worst efficiency among the methods considered here.

We note that the uEWMA charts rapidly detect large process changes. We numerically summarize the efficiency of the various methods in Table II at the clinically significant mean shift of $q=3.7$ which corresponds to a doubling in the odds of 30-day mortality. For instance the median time to signal for the uEWMA with log-likelihood ratio scores is around 56.4 days or 152 surgical procedures (at the rate of 2.7 patients per day) which equates to around 10 extra deaths given the overall mortality rate of 6.6 per cent.

The median run length results presented in Figure 5 and Table II depend on the patient arrival rate (and patient mix, survival model, etc.). We verified with simulation that the relative performance of the uEWMA and Biswas CUSUM is very similar at different patient arrival rates. The relative performance of the Sego CUSUM, however, does depend on the patient arrival rate. Although it is never better than the uEWMA and Biswas CUSUM, for low patient arrival rates the effect of the inherent 30-day delay is less important and vice versa for higher patient arrival rates.

7. Summary and discussion

The proposed updating EWMA (uEWMA) can be used for risk-adjusted monitoring of survival times. The uEWMA calculations maintain the ordering of patients and the method is thus appropriate for monitoring a process with an initial event (e.g. surgery) expected to drive the patient outcomes given in terms of the survival time. The uEWMA operates in continuous time updating patient scores with the latest information as time passes. The uEWMA can be based on any survival-time model and a variety of patient scores are possible, including scores based on the log-likelihood ratio and scores that have a more clinically relevant interpretation.

In terms of efficiency, we expect (so long as process changes are reasonably modeled as changes in the scale or location parameter of the survival-time model) that the uEWMA will signal process changes faster than methods such as Steiner *et al.* [7] that monitor only binary outcomes. In addition, due to continuous updating the uEWMA is shown here to signal more quickly than the Sego *et al.* [10] CUSUM that must wait until a set period of time (e.g. 30 days) after surgery to update the monitoring. This advantage would be even greater in situations where the follow-up time was longer, such as the kidney transplant example in Biswas and Kalbfleisch [11]. For the given example, the uEWMA chart with log-likelihood ratio scores is more efficient than the Biswas and Kalbfleisch [11] CUSUM. Using the uEWMA with the observed minus expected deaths scores or the survival model summary scores rather than the log-likelihood ratio scores results in minor loss of efficiency.

The uEWMA statistic, given in (1), is a weighted average score and thus provides a local estimate of the average score. As a result, unlike methods based on the CUSUM, generally EWMA control charts are easier to understand and interpret. This can be a big advantage when trying to promote monitoring procedures to medical clinicians. Determining the run length properties of the uEWMA requires simulation, however this is only needed to set up the chart and not during ongoing application.

In the presented cardiac surgery example we set the maximum follow-up time for all patients at 30 days. That is we assume that survival after 30 days may not have been driven mostly by the surgical quality. As seen in Section 3, competing risks can also be easily incorporated and lead to another possible reason for a censored observation. The time to a competing risk can be variable. In this situation the uEWMA makes use of all the relevant information in an appropriate way, e.g. patients having a competing risk event (say discharge from hospital) are included and scores are updated up until the time of the competing risk event.

In this paper we propose estimating the survival model parameters using 'in-control' data from phase I. In this way the monitoring scheme is designed to look for changes in the current process state. An alternative approach is to set the survival model parameters based on a standard process state as determined from the published literature or estimated from data aggregated across a number of centers (see for example [11]). With this approach a signal no longer suggests a process change; rather a signal shows that enough evidence has been collected to conclude that the performance of the monitored center is different from the established standard. This different interpretation should lead to a different reaction to chart signals with the two approaches.

In monitoring medical (and other) processes, it can be desirable to stratify results, for example by surgeon/hospital etc. [7]. This can make the monitoring more sensitive to local changes. Grigg *et al.* [28] discuss the issue of local and relative changes in a specific context. However, with any monitoring method, we need to be careful that with the simultaneous use of multiple control charts the overall median time to a false alarm does not become very short. Too frequent false alarms will lead to a monitoring method that practitioners will simply ignore.

The current methodology is relevant for adverse events due to a singly indexed event, e.g. surgery; however, it would be clinically useful to extend the methodology to repeated/regular events, e.g. maintenance of central venous catheter (CVC) lines, where the time to infection is relevant. Finally, with the updating scores it would also be possible to allow for time-dependent covariates if this turns out to be clinically reasonable and the covariates are not affected by the surgery.

References

1. Shewhart WA. *Economic Control of Quality of Manufacturer Product*. D Van Nostrand: New York, 1931 (Republished in 1981 by the American Society for Quality, Milwaukee, WI).
2. De Leval MR, François K, Bull C, Brawn WB, Spiegelhalter D. Analysis of a cluster of surgical failures. *The Journal of Thoracic and Cardiovascular Surgery* 1994; **107**:914–924.
3. Steiner SH, Cook R, Farewell V. Monitoring paired binary surgical outcomes using cumulative sum charts. *Statistics in Medicine* 1999; **18**:69–86.
4. Lovegrove J, Valencia O, Treasure T, Sherlaw-Johnson C, Gallivan S. Monitoring the results of cardiac surgery by variable life-adjusted display. *Lancet* 1997; **18**(350(9085)):1128–1130.
5. Lovegrove J, Sherlaw-Johnson C, Valencia O, Treasure T, Gallivan S. Monitoring the performance of cardiac surgeons. *Journal of the Operational Research Society* 1999; **50**:684–689.
6. Poloniecki J, Valencia O, Littlejohns P. Cumulative risk adjusted mortality chart for detecting changes in death rate: observational study of heart surgery. *British Medical Journal* 1998; **316**:1697–1700.
7. Steiner SH, Cook RJ, Farewell VT, Treasure T. Monitoring surgical performance using risk-adjusted cumulative sum charts. *Biostatistics* 2000; **1**:441–452.
8. Grigg O, Farewell VT. An overview of risk-adjusted charts. *Journal of the Royal Statistical Society, Series A* 2004; **167**:523–539.
9. Woodall WH. The use of control charts in health-care and public-health surveillance (with Discussion). *Journal of Quality Technology* 2006; **3**:89–104.
10. Sego LH, Reynolds Jr MR, Woodall WH. Risk adjusted monitoring of survival times. *Statistics in Medicine* 2009; **28**:1386–1401.
11. Biswas P, Kalbfleisch JD. A risk-adjusted CUSUM in continuous time based on the Cox model. *Statistics in Medicine* 2008; **27**:3382–3406.
12. Lucas JM, Saccucci MS. Exponentially weighted moving average control schemes: properties and enhancements. *Technometrics* 1990; **32**:1–12.
13. Roberts SW. Control chart tests based on a geometric moving averages. *Technometrics* 1959; **1**:239–250.
14. Morton A, Whitby M, McLaws ML, Dobson A, McElwain S, Looke D, Stackelroth J, Sartor A. The application of statistical process control charts to the detection and monitoring of hospital acquired infections. *Journal of Quality Clinical Practice* 2001; **21**:112–117.
15. Spliid H. Monitoring medical procedures by exponential smoothing. *Statistics in Medicine* 2007; **26**:124–138.
16. Grigg O, Spiegelhalter D. A simple risk-adjusted exponentially weighted moving average. *Journal of the American Statistical Association* 2007; **102**:140–152.
17. Box G, Jenkins G, Reinsel GC. *Time Series Analysis: Forecasting and Control* (4th edn). Holden-Day: San Francisco, 2008.
18. Crowder SA, Hamilton MD. An EWMA for monitoring a process standard deviation. *Journal of Quality Technology* 1992; **24**:12–21.
19. Woodall WH, Mahmoud MA. The inertial properties of quality control charts. *Technometrics* 2005; **47**:425–436.
20. Klein JP, Moeschberger ML. *Survival Analysis: Techniques for Censored and Truncated Data*. Springer: New York, NY, 1997.
21. Jones LA, Champ CW, Rigdon SE. The performance of exponentially weighted moving average control charts with estimated parameters. *Technometrics* 2001; **43**:156–167.
22. Jensen WA, Jones-Farmer LA, Champ CW, Woodall WH. Effects of parameter estimation on control chart properties: a literature review. *Journal of Quality Technology* 2006; **38**:349–364.
23. Moustakides GV. Optimal stopping times for detecting changes in distributions. *The Annals of Statistics* 1986; **14**:1379–1387.
24. Steiner SH. EWMA control charts with time-varying control limits and fast initial response. *Journal of Quality Technology* 1999; **31**:75–86.
25. Parsonnet V, Dean D, Bernstein A. A method of uniform stratification of risks for evaluating the results of surgery in acquired adult heart disease. *Circulation* 1989; **79**:1–12.
26. Steiner SH, MacKay RJ. Monitoring processes with highly censored data. *Journal of Quality Technology* 2000; **32**:199–208.
27. Cook DA. The development of risk adjusted control charts and machine learning models to monitor the mortality rate of intensive care unit patients. *Ph.D. Thesis*, School of Information Technology and Electrical Engineering, The University of Queensland, 2003.
28. Grigg OE, Spiegelhalter DA, Jones HE. Local and marginal control charts applied to methicillin resistant *Staphylococcus aureus* bacteraemia reports in UK acute National Health Service trusts. *Journal of the Royal Statistical Society, Series A* 2009; **172**:49–66.