

Risk-Adjusted Performance Monitoring in Healthcare Quality Control

Li Zeng

1 Introduction

With the growing emphasis and concerns on quality of health care, performance monitoring of care providers has received much attention recently. Performance measures used as monitors are typically clinical outcomes, utilization of health services, and cost. By monitoring these measures continuously, changes in the performance of care providers can be detected promptly to avoid serious consequences as well as provide valuable information on the care delivery system for quality improvement.

One critical challenge in performance monitoring in medical contexts is the need to adjust for patient case mix, called *risk adjustment* (Iezzoni 1997). Unlike products in manufacturing processes which are relatively homogeneous in nature, patients vary a lot in their characteristics or risk factors, which may affect the performance monitors. For example, sicker patients tend to experience worse outcomes, even with excellent care, than their healthier counterparts. The affecting patient risk factors must be taken into account in the monitoring to fairly assess the performance of care providers. Performance monitoring with such considerations is referred to as risk-adjusted (RA) monitoring in the literature.

Two basic problems are involved in RA monitoring, as illustrated in Fig. 1: *establishing risk adjustment models*, which includes identifying the appropriate performance measures to monitor and associated patient risk factors, constructing statistical models that characterize the dependency of the performance measures on the risk factors, and *change detection based on the established models*, which includes estimating baseline parameters of the risk adjustment models and detecting deviations from them. The former has been a focus of physicians and medical

L. Zeng (✉)

Department of Industrial and Systems Engineering,
Texas A&M University, College Station, TX 77843, USA
e-mail: lizeng@tamu.edu

quality researchers, while the later has attracted the attention of statisticians, including those industrial statisticians who are extending their statistical process control (SPC) research in industrial contexts to medical contexts. The objective of this chapter is to review the main developments on the two problems. A case study will also be provided to demonstrate the use of a powerful method, Bayesian approaches, for RA monitoring. It is worth mentioning that, unlike previous reviews of this topic which merely focus on the change detection problem, establishment of risk adjustment models is also considered in this chapter to provide a broader view of the RA monitoring problem which can help readers to understand the techniques for change detection better as well as enable the identification of potential collaboration opportunities between researchers in different areas.

2 Risk Adjustment Models

As the basis and first step in risk-adjusted monitoring, a statistical risk adjustment model must be constructed based on domain knowledge of the application of interest and historical data available. As shown in Fig. 1, a risk adjustment model consists of three components: performance measures to monitor, patient risk factors that may affect the performance measures, and statistical models that characterize the dependency of performance measures on the risk factors. Such models vary from one application to another. Typical examples and considerations in constructing these models will be introduced in this section.

2.1 Performance Measures and Patient Risk Factors

Risk adjustment models have been developed in many critical areas in health care in the past two decades, such as thoracic and cardiac surgeries (e.g., Brunelli et al. 2007; Daley et al. 2001; Krumholz et al. 1999; Pinna-Pintor et al. 2002; Shroyer et al. 2003; Sousa et al. 2008; Tu et al. 1995), public mental health (e.g., Hendryx et al. 1999; Hermann et al. 2007), home health care (e.g., Murtaugh et al. 2007),

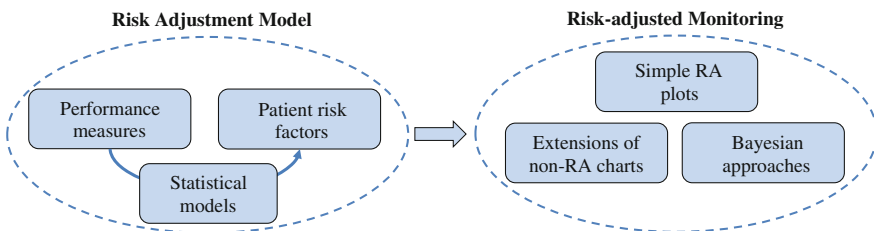


Fig. 1 The framework of risk-adjusted monitoring

Table 1 Examples of performance measures and risk factors in some applications

Application	Key performance measures	Important patient risk factors
Surgeries	30-day mortality survival time	Age, sex, race, previous myocardial infarction, cerebrovascular disease, renal failures requiring
	Presence of major complications	Dialysis
Public mental health	Inpatient length of stay inpatient cost patient satisfaction change in functioning	Age, illness severity, income, prior utilization of mental health services, amount of social support received
Home health care	Improvement in bathing stabilization in speech acute care hospitalization	Age, socioeconomic factors, prior service use, sensory status, diagnosis severity
Nursing home care	Decline in functional status worsening pressure ulcers mortality	Age, sex, medical conditions, physical restraints, decubiti at admission
General hospital care	Presence of readmission inpatient mortality occurrence of patient safety events	Age, sex, principal diagnosis, number of major chronic conditions and significant comorbidities

nursing home care (e.g., Mukamel and Spector 2000; Zimmerman 2003), and general hospital care (e.g., Benbassat and Taragin 2000; Forthman et al. 2010). Cardiac surgeries is the most popular area for such studies, a fact that is due largely to the motivation of well-publicized reports of surgical outcomes and cases where high rates of surgical complications remained undetected for an undue length of time (Steiner et al. 2000). Examples of key performance measures and important affecting risk factors in the above-mentioned applications are listed in Table 1.

The performance measures are typically clinical outcomes, especially adverse events, such as mortality and morbidities, service utilization, length of stay in intensive care unit, patient satisfaction, and cost that are commonly recognized indicators of quality of care in terms of its six main dimensions including effectiveness, efficiency, timeliness, patient-centeredness, equality, and safety (Reid 2005). The most widely used performance measure is mortality. The use of this measure dates back to the releasing of patient mortality records by the Health Care Financing Administration (HCFA) in the 1980s, which aroused much criticism then, but has stimulated discussions on how to measure quality of care (DesHarnais et al. 1991).

There are many affecting patient risk factors associated with a chosen performance measure, ranging from patients' demographic characteristics, diagnosis, severity of illness, medical conditions to socioeconomic status and social support received. To provide convenience in implementing risk adjustment in practice, the effects of different risk factors are often combined into a risk score such as the Parsonnet score and APACH score that have been widely used in cardiac surgeries and intensive care (Iezzoni 1997).

2.2 Statistical Models

With data of performance measures and patient risk factors, statistical models are built to characterize their relationship. The data are typically obtained from administrative claims database or patient medical records. As popular performance measures in health care fall into two categories, binary measures (e.g., death/survival, presence/absence of certain complications) and continuous measures (e.g., survival time following a surgery, length of hospital stay, cost), statistical models for the two types of measures have been developed in existing studies.

Specifically, the logistic regression model is commonly used for binary measures

$$y \sim \text{Bernoulli}(p)$$

$$\log \frac{p}{1-p} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p \quad (1)$$

where y is the performance measurement which takes value 1 or 0 corresponding to the occurrence or not of the concerned adverse event such as death, p is the probability of the occurrence of the event, which is the parameter of the Bernoulli distribution that generates y , x_1, \dots, x_p are the affecting patient risk factors, and β_1, \dots, β_p are the corresponding coefficients in the model which represent the effects of the risk factors on the performance measure.

There are two types of continuous performance measures, regular ones (e.g., cost), and time to event (e.g., survival time following a surgery). For the former, standard linear regression models are used

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \varepsilon \quad (2)$$

where ε is the error term assumed to follow a normal distribution. For the latter, different survival models are used such as the accelerated failure time model

$$\log y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + e \quad (3)$$

where the distribution of the error term e can take different forms defined on $[0, \infty)$, such as normal distribution, logistic distribution, and extreme value distribution leading to log-normal, log-logistic, and Weibull distributions for y . More complex models may also be used such as the Cox proportional hazard models.

In constructing risk adjustment models, two issues need to be considered: First, variable selection. Selection of significant variables, a basic task in statistical regression analysis, is especially critical in constructing risk adjustment models because a large pool of patient characteristics is typically available in healthcare databases which may contain a lot of irrelevant or redundant information. Moreover, models involving too many variables may also pose difficulties to data collection in practice in terms of time and cost. Simple methods, such as stepwise selection procedures, have been used in the existing studies for this purpose.

Second, dealing with multiple performance measures. Multiple performance measures normally exist to characterize the quality of care in different aspects. The measures may bear complex correlations which need to be taken into account in risk adjustment (DesHarnais et al. 1991). Despite a common recognition of this issue, however, no formal analysis has been done about it in the current literature than simply combining multiple performance measures into one single measure.

3 Risk-Adjusted Performance Monitoring

With the established risk adjustment model, the performance of care providers will be characterized by the parameters of the model, and then monitoring will be conducted to detect changes in the parameters. An example of data used in RA monitoring is shown in Fig. 2, where the upper panel displays a stream of performance measurements in cardiac surgeries, i.e., patient mortality status within 30 days after the surgery (D/S denoting death/survival), and the lower panel shows the associated Parsonnet scores of patients, which indicate their preoperative risk of death.

Before introducing the techniques for RA monitoring, definitions of important concepts on performance monitoring are provided below to facilitate understanding of those techniques:

Phase I versus Phase II monitoring These are the two basic types of monitoring considered in SPC research. Phase I monitoring, also referred to as retrospective monitoring, aims to detect changes in performance during a fixed time period, while Phase II monitoring, also referred to as prospective monitoring, aims to detect deviation of performance from a baseline whenever a data point becomes available. In Phase I monitoring, the data sequence is fixed, and the goal is to

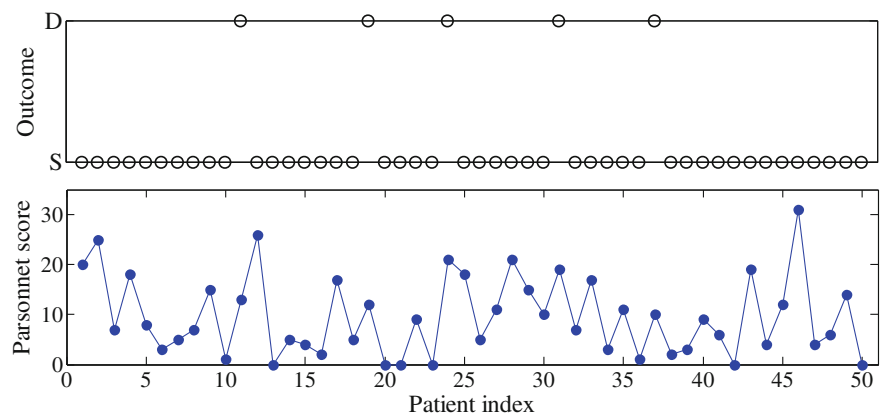


Fig. 2 A typical data set used in risk-adjusted performance monitoring

identify change points in the data as accurately as possible. The baseline is normally unknown, and needs to be estimated in Phase I monitoring. In Phase II monitoring, the data are obtained sequentially, and the goal is to capture a change as soon as possible. A baseline estimated from historical data is typically available, and a change is defined to be a deviation from the baseline. There is also a hybrid type of monitoring, called self-starting monitoring, which takes the online detection scheme in Phase II monitoring, but without a known baseline as in Phase I monitoring.

Grouped monitoring versus continuous monitoring In grouped monitoring, patients are divided into groups of similar sizes, and some aggregate summaries of each group will be monitored. This is consistent with the concept of “subgroup” in industrial SPC research. An alternative way is to monitor each patient individually, called continuous monitoring in the literature. Essentially, the continuous monitoring scheme represents 100 % inspection, which is supposed to be a better scheme for performance monitoring in medical contexts where prompt detection of changes is particularly desired (Woodall 2006).

Grigg and Farewell (2004a, b), Woodall (2006), and Cook et al. (2008) give excellent reviews on methods and techniques for RA performance monitoring. In this chapter, a brief review of the basics and advantages/disadvantages of popular RA monitoring techniques will be given in a unified way, with an emphasis on techniques developed after those reviews and, particularly, Bayesian approaches. These methods can be divided into three categories: simple RA plots, extensions of non-RA SPC control charts, and Bayesian approaches. All the techniques in the first two categories focus on Phase II monitoring except the last one in the second category, while Bayesian approaches can be used for both Phase I and Phase II monitoring.

3.1 *Simple Risk-Adjusted Plots*

Simple plots of the cumulative difference between observed and expected outcomes have been used to signal changes in surgical performance

$$C_t = C_{t-1} + (y_t - p_t)$$

where C_t is the statistic at time t , y_t is the observed outcome (death/survival), and p_t is the baseline mortality probability of patient t . Obviously, if there is a sustained change in the performance of care providers, an increasing/decreasing trend can be seen from the plot. Such methods include the observed-expected (O-E) plot (Poloniecki 1998) and the variable life-adjusted (VLAD) plot (Lovegrove et al. 1997). While they are easy to implement and understand by healthcare practitioners, the statistical properties of the statistic monitored are not clear and thus it is difficult to set up control limits.

3.2 Extensions of Non-risk-Adjusted SPC Control Charts

As SPC is a well-studied area in other contexts especially industrial applications, various non-risk-adjusted control charts are available in the literature. These techniques have been extended to medical contexts by incorporating risk adjustment. Popular extensions of these techniques will be briefly introduced as follows.

3.2.1 Risk-Adjusted p -Chart

p -chart is a basic SPC technique to monitor binary data such as defectiveness or non-defectiveness of products in industrial process control where defective rate is an important concern. To apply this technique in Phase II monitoring, subgroups of data need to be collected, and a normal distribution is assumed when the sample size of subgroups is adequately large

$$\hat{p}_i = \frac{\sum_{t=1}^{n_i} y_{it}}{n_i} \sim N\left(p_0, \frac{p_0(1-p_0)}{n_i}\right)$$

where n_i is the sample size of subgroup i , \hat{p}_i is the corresponding average defective rate, y_{it} is the measurement of product t in subgroup i , and p_0 is the baseline defective rate estimated from historical data. 3-sigma control limits can be obtained based on this distribution.

To extend this method to medical contexts, patients in consecutive time periods of same length, e.g., 6 months, are grouped, and the distribution used in the monitoring becomes

$$\hat{p}_i = \frac{\sum_{t=1}^{n_i} y_{it}}{n_i} \sim N\left(\frac{1}{n_i} \sum_{t=1}^{n_i} p_{0ti}, \frac{1}{n_i} \sum_{t=1}^{n_i} p_{0ti}(1-p_{0ti})\right)$$

where p_{0ti} is the baseline mortality probability of patient t in group i . The simple idea here is to represent the mortality probability of each subgroup by averaging the mortality probabilities of patients in that group, which is reasonable when the number of patients in the group is large enough. Cockings et al. (2006) and Cook et al. (2003) use such charts to monitor mortality in intensive care.

The RA p -chart bears the merits of the simple plots, that is, easiness in implementation and interpretation, and also provides a convenient way to set up control limits. However, the need for grouping patients during considerably long periods may lead to delay in capturing changes in performance, which makes it not so appealing as techniques designed for continuous monitoring.

3.2.2 Risk-Adjusted Set Method

The set method monitors the time between events of interest (e.g., death) by counting the number of events (e.g., survival) between. Specifically, letting C_t be the current set number, i.e., count of events following the occurrence of an interested event, $C_t = C_{t-1} + 1$, that is, this statistic will increase by 1 if the t th observation is not the interested event. This continues until an interested event occurs, and then the set number will be reset to 0. An alarm is signaled when $C_t \leq T$ happens n times, where (T, n) is a pair of thresholds determined through simulation.

An extension of this method to incorporate risk adjustment has been proposed by Grigg and Farewell (2004a, b). The basic idea is to weigh each event by the baseline mortality probability of the patient. Specifically, the set number will be calculated by

$$C_t = C_{t-1} + \frac{p_{0t}}{\bar{p}_0}$$

where p_{0t} is the baseline mortality probability of patient t , and \bar{p}_0 is the average baseline mortality probability of all patients, which can also be termed as the baseline mortality probability of an “average” patient. Here the average patient is used as a benchmark to assess the normality of each observation, and patients with a higher baseline mortality probability than the average patient can be assigned a higher weight.

The set method also provides a graphical representation, called grass plot, to assist decision making. The drawbacks of this method lie in the complexity in determining the paired thresholds and inference based on the time between events rather than individual observations, which may cause delay in change detection.

3.2.3 Risk-Adjusted CUSUM Chart

Cumulative sum (CUSUM) control charts is a popular SPC technique due to their optimal properties. In the general setting, such charts aim to test the following hypotheses:

$$H_0 : \theta = \theta_0 \quad H_1 : \theta = \theta_1$$

where θ denotes the parameter of the risk adjustment model, θ_0 is the baseline which is typically known, and θ_1 is a hypothesized value of interest. The following statistic is monitored

$$C_t = \max(0, C_{t-1} + W_t)$$

where W_t is the CUSUM score assigned to the t th observation. A control limit H will be found through simulation to achieve a specified in-control average run length (ARL_0), and an alarm is signaled when $C_t > H$.

The CUSUM score is given by the log-likelihood ratio of the two hypotheses

$$W_t = \log \left(\frac{L(\theta_1|y_t)}{L(\theta_0|y_t)} \right) \quad (4)$$

where y_t is the t th observation, and $L(\theta|y_t)$ is the likelihood function of the risk adjustment model. For example, for binary data following a Bernoulli distribution with parameter $\theta = p$, the likelihood function is

$$L(\theta|y_t) = \theta^{y_t} (1 - \theta)^{1-y_t}$$

CUSUM charts based on the above likelihood have been used widely to monitor defective rate of products in industrial processes.

For binary performance measures in healthcare which follow the logistic regression model in (1), the likelihood function is

$$L(\theta|y_t) = \left(\frac{Rp_{0t}}{1 - p_{0t} + Rp_{0t}} \right)^{y_t} \left(\frac{1 - p_{0t}}{1 - p_{0t} + Rp_{0t}} \right)^{1-y_t}$$

where the parameter of interest $\theta = R$, R is the odds ratio, and p_{0t} is the baseline mortality probability of patient t . The resulting CUSUM charts are the risk-adjusted version of CUSUM since p_{0t} , which depends on patient risk factors, is taken into account in the statistic.

When the performance measure is time to event such as survival times, and the accelerated failure time model in (3) is used, the likelihood function depends on the assumed distribution for the survival times. For example, if the Weibull distribution is used, the likelihood function is

$$L(\theta|y_t) = \frac{\alpha}{\theta} \left(\frac{y_t}{\theta} \right)^{\alpha-1} \exp \left[- \left(\frac{y_t}{\theta} \right)^\alpha \right]$$

where $\theta = \lambda$, λ is the scale parameter, and α is the shape parameter which is often assumed to be fixed and can be estimated from historical data. A special issue in dealing with survival data is censoring where the observation y_t is either the survival time of patient t or the censoring time. In this case, a likelihood function taking into account censoring will be used.

Risk-adjusted CUSUM charts for binary performance measures were first proposed by Steiner et al. (2000) in monitoring 30-day mortality in cardiac surgeries, and then applied in other applications such as liver transplant to monitor 1-year mortality (Leandro et al. 2005) and coronary artery bypass surgeries to monitor adverse outcomes (Novick et al. 2006). RA CUSUM charts for time to event were developed by Sego et al. (2009), Gandy et al. (2010), and Biswas and Kalbfleisch

(2008), who use different models for the survival time. These charts, like their non-risk-adjusted counterparts in industrial contexts, are powerful in detecting small changes in performance, but their use is limited by the perceived difficulty of interpretation by healthcare practitioners (Cook et al. 2011; Pilcher et al. 2010).

3.2.4 Risk-Adjusted EWMA Chart

Like CUSUM charts, the exponentially weighted moving average (EWMA) charts are a popular and widely used SPC technique. The statistic monitored in these charts takes the following form:

$$C_t = \gamma S_t + (1 - \gamma)C_{t-1}$$

where S_t is the EWMA score assigned to the t th observation and $0 < \gamma \leq 1$ is a smoothing constant. Essentially, the statistic is a linear combination of all the observations with higher weights assigned to recent observations. With the linearity in the statistic, its distribution can be obtained analytically and, consequently, control limits can be specified based on that.

There are different definitions for the EWMA score depending on the types of data monitored. For binary performance measures, S_t can be the baseline mortality probability or the difference in the observed and baseline mortality probability (Cook et al. 2008; Cook et al. 2011). For time to event performance measures, S_t can be the likelihood ratio scores as used in RA CUSUM charts (Steiner and Jones 2010).

The RA EWMA charts have similar performance to CUSUM charts in detecting small changes. Its main advantage over CUSUM charts lies in its intuitive interpretation as the EWMA statistic can be viewed as an estimate of the current level of the process. Moreover, the influence of previous observations is removed in the statistic gradually by adjusting the weights rather than resetting the statistic as CUSUM does, which is a more natural way to conduct monitoring and easier to accept by healthcare practitioners (Cook et al. 2011).

3.2.5 Likelihood Ratio Test for Phase I Monitoring

The above control charts are all designed for Phase II monitoring. Kamran et al. (2012) propose a control chart based on likelihood ratio test for Phase I monitoring. This method is built on the change-point setting which tests the following hypotheses:

$$\begin{aligned} H_0: y_i &\sim \text{LG}(y_i|\theta_0) & i = 1, \dots, m \\ H_1: y_i &\sim \begin{cases} \text{LG}(y_i|\theta_0) & i = 1, \dots, K \\ \text{LG}(y_i|\theta_1) & i = K + 1, \dots, m \end{cases} \end{aligned} \quad (4)$$

where $\text{LG}(\cdot|\theta)$ is the logistic regression model with parameter θ , m is the total number of observations available, and K , $1 \leq K \leq m - 1$, is the change point at

which the model parameter changes from θ_0 to θ_1 , $\theta_1 \neq \theta_0$. A likelihood ratio statistic can then be constructed as

$$\Lambda(\tau) = \log \left(\frac{L(\hat{\theta}_0^1, \hat{\theta}_1^1 | y_1, \dots, y_m)}{L(\hat{\theta}_0^0 | y_1, \dots, y_m)} \right)$$

where $\hat{\theta}_0^1$, $\hat{\theta}_1^1$ and $\hat{\theta}_0^0$ are maximum likelihood estimates of parameters under the two hypotheses. Control limits will be determined through simulation.

3.3 Bayesian Approaches for RA Monitoring

Bayesian approaches have been used for process monitoring and change detection in various applications. Recently, such approaches are developed for different RA monitoring problems, including Phase I monitoring (Assareh et al. 2011a, b; Assareh and Mengersen 2012), estimating the location where change in performance occurs (Assareh et al. 2011c), and self-starting performance monitoring (Zeng and Zhou 2011). This section will first present the Bayesian framework for change detection in RA monitoring, and then its applications in different specific problems. Its advantages and disadvantages over the abovementioned non-Bayesian techniques will be summarized in the end.

3.3.1 Bayesian Framework for Change Detection

Assume the data monitored follow a change-point model as in (4)

$$y_i \sim \text{CP}(y_i | \theta_0, K, \theta_1) = \begin{cases} \text{LG}(y_i | \theta_0) & i = 1, \dots, K \\ \text{LG}(y_i | \theta_1) & i = K + 1, \dots, m \end{cases}$$

In the Bayesian framework, the unknown change point K is treated as a parameter of the change-point model $\text{CP}(\cdot)$. Here this model is characterized by three sets of parameters, the pre-change parameter θ_0 , the change point K , and the post-change parameter θ_1 . Correspondingly, any inference regarding this model relates to finding the posterior distribution of these parameters:

$$p(\theta_0, K, \theta_1 | y_1, \dots, y_m) = \pi(\theta_0, K, \theta_1) \cdot f(y_1, \dots, y_m | \theta_0, K, \theta_1) \quad (5)$$

where $p(\cdot | y_1, \dots, y_m)$ is the posterior, $\pi(\cdot)$ is the prior, and $f(\cdot)$ is the sampling density as follows:

$$f(y_1, \dots, y_m | \theta_0, K, \theta_1) = \prod_{i=1}^K \text{LG}(y_i | \theta_0) \cdot \prod_{i=K+1}^m \text{LG}(y_i | \theta_1)$$

Samples from the posterior distribution can be obtained through Markov chain Monte Carlo (MCMC) algorithms, and summaries of these samples will be used for decision making in performance monitoring.

One critical step in obtaining the posterior samples is to specify the priors. Assuming that the three sets of parameters are independent, their priors can be specified separately, that is,

$$\pi(\theta_0, K, \theta_1) = \pi(\theta_0) \cdot \pi(K) \cdot \pi(\theta_1)$$

For θ_0 and θ_1 , specifying their priors is equivalent to specifying the prior for the logistic regression model in (1), where the parameter $\theta = [\beta_0, \dots, \beta_p]'$. This problem has been considered in many studies, and appropriate choices depend on the availability of prior information such as historical data and expert knowledge. When there is prior information, the prior of θ can be either estimated from historical data or elicited from expert knowledge; otherwise regular priors such as flat priors, normal priors, and conjugate priors can be used. Zeng and Zhou (2011) propose ideas on specifying priors for θ in both cases in RA monitoring. For the change point K , a uniform prior on $\{1, 2, \dots, m-1\}$ is commonly used.

3.3.2 Bayesian Estimation of Change Points

As suggested by Assareh et al. (2011a, c), Bayesian approaches can be used in conjunction with the non-Bayesian control charts such as RA CUSUM charts to estimate the location of the change point when a change is detected using those charts. Summaries, such as mean, median, and mode of the posterior samples can be used as estimates of the change point.

3.3.3 Bayesian Phase I Monitoring

The central task of Phase I monitoring is to determine if there is any change point in the performance during the studied time period. This can be formulated as testing the following hypotheses regarding the parameter K :

$$H_0 : K \leq \tau_L \text{ or } K \geq \tau_U \quad H_1 : \tau_L < K < \tau_U \quad (6)$$

where τ_L and τ_U , $\tau_L > 1$, $\tau_U < m-1$ are the specified lower and upper bounds. Such bounds are needed in decision making because very little evidence could exist to support K being at the very beginning or end of the data sequence.

This problem can be solved by using the Bayes factor (BF), which is a popular Bayesian tool for model comparison. Essentially, the Bayes factor compares the marginal likelihoods under the two hypotheses to determine the plausibility of one against another

$$\text{BF}(H_1 : H_0) = \frac{P(y_1, \dots, y_m | H_1)}{P(y_1, \dots, y_m | H_0)}$$

A value of BF larger than a chosen threshold, η , $\eta > 1$, means that H_1 is more strongly supported by the data. BFs can be obtained from the posterior samples (Zeng and Zhou 2011).

In practice, there are possibilities that multiple change points may exist in the data. A simple binary segmentation strategy can be applied, that is, we first try to capture one change point in the dataset; if a change point is detected, the data will be broken into two segments by the identified change point, and then the procedure will be applied to each segment to capture one change point in that segment. This repeats until all the change points are identified. A more advanced way is to explicitly represent the number of change points and the locations of change points as random variables in the change-point model, and then find posterior distribution of the expanded parameter space.

3.3.4 Bayesian Phase II Monitoring

In Phase II monitoring, the number of observations (i.e., m) increases over time. For each value of m , the Bayes factor will be calculated and decision will be made on whether some change has occurred. If not, the monitoring will continue; otherwise, we will stop and estimate the change point. Since a baseline is normally available for Phase II monitoring, the pre-change parameter θ_0 in the change-point model will be known and, consequently, the posterior distribution in (5) contains only two parameters, θ_1 and K . Obviously, this scheme can also be used for self-starting monitoring where the baseline is unknown.

3.3.5 Advantages/Disadvantages of Bayesian Approaches

Compared to non-Bayesian RA monitoring techniques, Bayesian approaches have the following advantageous features:

1. *Simple, generic, and versatile framework*: The Bayesian framework based on posterior inference provides a simple and generic way to conduct risk-adjusted monitoring for different types of risk adjustment models. This framework can also be easily adapted to solve different types of monitoring problems.
2. *Intuitive interpretation*: The Bayes factor bears an intuitive interpretation as evidence of the plausibility of one hypothesis against another. Moreover, as will be shown in the case study, the sampled posterior distribution of parameters provides an intuitive graphical representation of possible locations of the change point and associated uncertainty, which will be a valuable tool for medical practitioners in decision making.

3. *Use of priors:* As a defining feature of Bayesian statistics, the use of priors provides a way to incorporate domain knowledge of physicians and other medical professionals in the inference. This fits very well the medical contexts where expert knowledge is very critical.

The main drawback of Bayesian approaches lies in two aspects: the computation load in generating posterior distributions and computing Bayes factors, and the efforts needed for specifying the priors. The former, however, is not a significant challenge with the readily available MCMC algorithms. For the latter, prior setting in medical applications has been studied by many researchers (e.g., Chaloner and Duncan 1983; Chen et al. 2008). Those results need to be adapted to specific applications for RA monitoring.

4 Case Study

This section presents a case study to demonstrate Bayesian approaches for different RA monitoring problems. The data set used in this study is from a UK center for cardiac surgery, part of which is shown in Fig. 2. It contains information on operations during 1992–1998, including time of an operation, surgeon performing the operation, Patient Parsonnet score, and 30-day mortality following the operation. This data set has been used in many studies on RA monitoring (e.g., Steiner et al. 2000; Kamran et al. 2012). In this study, 1701 observations from a single surgeon will be used in the analysis, as displayed in Fig. 3, where the red dots indicate deaths. The logistic model for the data is

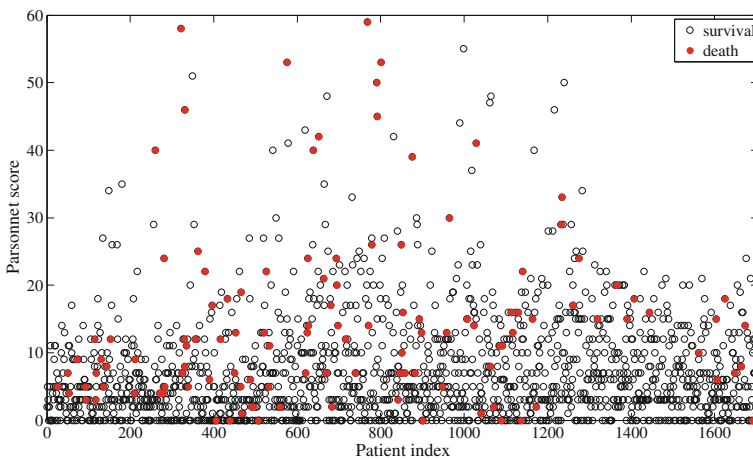


Fig. 3 The data used in the case study

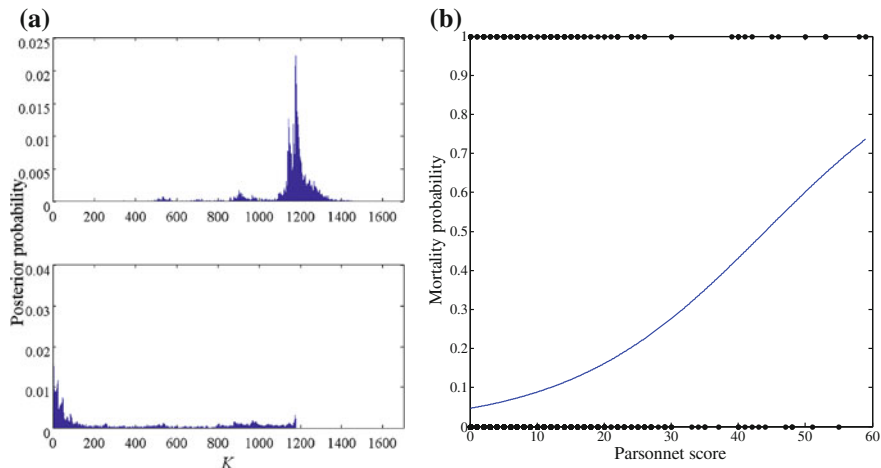


Fig. 4 Results in Phase I monitoring: posterior distribution of K (a) and fitted model (b)

$$\log \frac{p}{1-p} = \beta_0 + \beta_1 x$$

where x is the Parsonnet score and the parameter of this model is $\theta = [\beta_0, \beta_1]'$.

First, Phase I monitoring is conducted on the data. A vague normal prior $\pi(\beta_1) = N(0, 10^2)$ is used for parameter β_1 , and a flat prior $\pi(\beta_0) = \text{Uniform}(0, 0.5)$ is used for β_0 . Here β_0 must be positive because a higher risk score tends to lead to a higher mortality probability. The posterior samples of the change point K are obtained through slice sampler, a convenient MCMC algorithm. The empirical distribution of these samples is shown in the upper panel of Fig. 4a. A large number of samples concentrate on a small area around the 1200th patient, a sign that a change point may exist in the data. This is consistent with a rough observation on the raw data in Fig. 3. Since the evidence of change is very strong, the calculated Bayes factor is very large (>100) for any reasonable specification of τ_L and τ_U in (6). Therefore, we conclude that there is a change point in the data. The location of the change point is then estimated using the mode of the posterior of K , which is 1175. To examine if there is any change point before this one, the procedure is applied again to data of the first–1175 patients. The resulting posterior distribution of K is shown in the lower panel of Fig. 4a. The corresponding Bayes factor is very small (<1) for reasonable settings of τ_L and τ_U , meaning that there is no further change point.

Second, since it is determined that there is no change point during the first 1175 observations, these data are used to estimate the baseline parameters of the logistic regression model. Bayesian approach is used to conduct the estimation, which yields point estimates

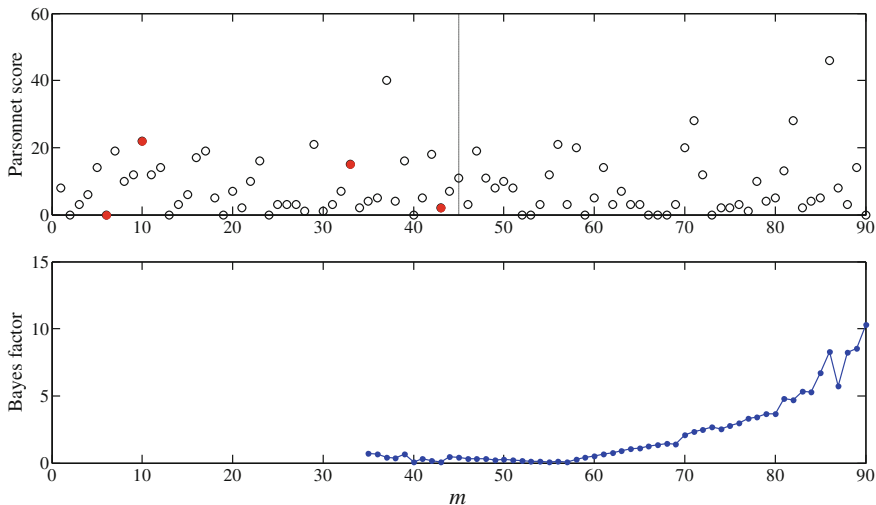


Fig. 5 Results in Phase II monitoring: data (*upper*) and resulting Bayes factors (*lower*)

$$\hat{\beta}_0 = -3.02, \quad \hat{\beta}_1 = 0.0686$$

The fitted logistic model is shown in Fig. 4b where the dots denote observed outcomes.

Finally, Phase II monitoring is conducted based on the baseline estimated in the Phase I analysis. For simplicity, only a segment of data in Fig. 3 is used. $\tau_L = 15$, $\tau_U = m - 15$, and the monitoring starts when $m = 35$. The data and the resulting Bayes factors are shown in Fig. 5, where the dashed line in the upper panel indicates the location of the change point identified in Phase I analysis. We can see that as m increases, the evidence of change becomes stronger which is manifested clearly by the increasing trend of Bayes factors.

5 Summary and Discussion

Risk-adjusted performance monitoring is a critical research area in healthcare quality control and has received much attention in recent years. Many studies have been done on this topic in different applications and for different purposes. This chapter gives an overview of the existing studies on RA monitoring, encompassing the basic elements of risk adjustment models and popular methods for change detection based on those models. A case study is provided to demonstrate the use of Bayesian approaches for different problems in RA monitoring.

Overall, this topic is an underdeveloped area, and there are many opportunities for future research. One potential direction is variable selection in the construction

of risk adjustment models, which is challenged by the existence of large amounts of patient characteristics data, as is often the case in statistical analysis in medical contexts. Powerful statistical/data mining techniques need to be applied to select significant patient risk factors and thus identify a parsimonious risk adjustment model which is the foundation for performance monitoring. Another direction is extending the current RA monitoring methods to more complex types of data than the univariate binary or continuous data. For example, as multiple performance measures are typically needed to characterize quality of care, a simultaneous monitoring scheme on those measures is desired. This can be solved by extending the multivariate control charting methods that have been well studied in industrial contexts to incorporate risk adjustment. Finally, as demonstrated in the case study, Bayesian approaches have great potential to be used for RA performance monitoring in health care. Efforts are needed to develop such approaches for different specific applications.

References

- Assareh, H., Smith, I., & Mengersen, K. (2011a). Bayesian change point detection in monitoring cardiac surgery outcomes. *Quality Management in Health Care*, 20(3), 207–222.
- Assareh, H., Smith, I., & Mengersen, K. (2011b). Change point detection in risk adjusted control charts. *Statistical Methods for Medical Research* (in press).
- Assareh, H., Smith, I., & Mengersen, K. (2011c). Bayesian estimation of the time of a linear trend in risk-adjusted control charts. *International Journal of Computer Science*, 38(4), 409–417.
- Assareh, H., & Mengersen, K. (2012). Change point estimation in monitoring survival time. *PLoS One*, 7(3), 1:10.
- Benbassat, J., & Taragin, M. (2000). Hospital readmissions as a measure of quality of health care. *Archives of Internal Medicine*, 160, 1074–1081.
- Biswas, P., & Kalbfleisch, J. D. (2008). A risk-adjusted CUSUM in continuous time based on cox model. *Statistics in Medicine*, 27, 3382–3406.
- Brunelli, A., Morgan-Hughes, N. J., Refai, M., Salati, M., Sabbatini, A., & Rocco, G. (2007). Risk-adjusted morbidity and mortality models to compare the performance of two units after major lung resections. *The Journal of Thoracic and Cardiovascular Surgery*, 133(1), 88–96.
- Chaloner, K., & Duncan, G. T. (1983). Assessment of A beta prior distribution: PM elicitation. *The Statistician*, 27, 174–180.
- Chen, M., Ibrahim, J. G., & Kim, S. (2008). Properties and implementation of Jeffreys's prior in binomial regression models. *Journal of the American Statistical Association*, 103, 1659–1664.
- Cocking, J. G. L., Cook, D. A., & Iqbal, R. K. (2006). Process monitoring in intensive care with the use of cumulative expected minus observed mortality and risk-adjusted p charts. *Critical Care*, 10(1), R28.
- 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, 469–474.
- Cook, D. A., Stefan, S. H., Cook, R. J., Farewell, V. T., & Morton, A. P. (2003). Monitoring the evolutionary process of quality: Risk-adjusted charting to track outcomes in intensive care. *Critical Care Medicine*, 31(6), 1676–1682.
- Cook, D. A., Duke, G., Hart, G. K., Pilcher, D., & Mullany, D. (2008). Review of the application of risk-adjusted charts to analyse mortality outcomes in critical care. *Critical Care and Resuscitation*, 10(3), 239–251.

- Daley, J., Henderson, W. G., & Khuri, S. F. (2001). Risk-adjusted surgical outcomes. *Annual Review of Medicine*, 52, 275–287.
- DesHarnais, S., McMahon, L. F., & Wroblewski, R. (1991). Measuring outcomes of hospital care using multiple risk-adjusted indexes. *Health Services Research*, 26(4), 425–445.
- Forthman, M. T., Gold, R. S., Dove, H. G., & Henderson, R. D. (2010). Risk-adjusted indices for measuring the quality of inpatient care. *Quality Management in Health Care*, 19(3), 265–277.
- Gandy, A., Kvaløy, J. T., Bottle, A., & Zhou, F. (2010). Risk-adjusted monitoring of time to event. *Biometrika*, 97(2), 375–388.
- Grigg, O., & Farewell, V. (2004a). An overview of risk-adjusted charts. *Journal of the Royal Statistical Society: Series A*, 167, 523–539.
- Grigg, O., & Farewell, V. T. (2004b). A risk-adjusted sets method for monitoring adverse medical outcomes. *Statistics in Medicine*, 23, 1593–1602.
- Hendryx, M. S., Dyck, D. G., & Srebnik, D. (1999). Risk-adjusted outcome models for public mental health outpatient programs. *Health Services Research*, 34(1), 171–195.
- Hermann, R. C., Rollins, C. K., & Chan, J. A. (2007). Risk-adjusting outcomes of mental health and substance-related care: A review of the literature. *Harvard Review of Psychiatry*, 15(2), 52–69.
- Iezzoni, L. I. (ed). 1997. *Risk Adjustment for Measuring Healthcare Outcomes*, 3rd edn. Chicago: Healthcare Administration Press.
- Kamran, P., 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.
- Krumholz, H. M., Chen, J., Wang, Y., Radford, M. J., Chen, Y.-T., & Marciniak, T. A. (1999). Comparing AMI mortality among hospitals in patients 65 years of age and older. *Circulation*, 99(23), 2986–2992.
- Leandro, G., Rolando, N., & Gallus, G. (2005). Monitoring surgical and medical outcomes: The Bernoulli cumulative SUM chart. A novel application to assess clinical interventions. *Postgraduate Medical Journal*, 81, 647–652.
- Lovegrove, J., Valencia, O., Treasure, T., Sherlaw-Johnson, C., & Gallivan, S. (1997). Monitoring the results of cardiac surgery by variable life-adjusted display. *The Lancet*, 350(18), 1128–1130.
- Mukamel, D. B., & Spector, W. D. (2000). Nursing home costs and risk-adjusted outcome measures of quality. *Medical Care*, 38(1), 78–89.
- Murtaugh, C. M., Peng, T., Aykan, H., & Maduro, G. (2007). Risk adjustment and public reporting on home health care. *Health Care Financing Review*, 28(3), 77–94.
- Novick, R. J., Fox, S. A., Stitt, L. W., Forbes, T. L., & Steiner, S. (2006). Direct comparison of risk-adjusted and non-risk-adjusted CUSUM analyses of coronary artery bypass surgery outcomes. *The Journal of Thoracic and Cardiovascular Surgery*, 132(2), 386–391.
- Pilcher, D. V., Hoffman, T., Thomas, C., Ernest, D., & Hart, G. K. (2010). Risk-adjusted continuous outcome monitoring with an EWMA chart: Could it have detected excess mortality among intensive care patients at Bundaberg base hospital? *Critical Care and Resuscitation*, 12(1), 36:41.
- Pinna-Pintor, P., Bobbio, M., Colangelo, S., Veglia, F., Giammaria, M., Cuni, D., et al. (2002). Inaccuracy of four coronary surgery risk-adjusted models to predict mortality in individual patients. *European Journal of Cardio-Thoracic Surgery*, 21, 199–204.
- Poloniecki, J., Valencia, O., & Littlejohns, P. (1998). Cumulative risk adjusted mortality chart for detecting changes in death rate: observational study of heart surgery. *British Medical Journal*, 316, 1697–1700.
- Reid, P. P., Compton, D., Grossman, J. H., Fanjiang, G., & Committee on Engineering and the Health Care System. (2005). *Building a better delivery system: A new engineering/health care partnership*. Washington, D.C.: National Academy Press.
- Sego, L. H., Reynolds, M. R., & Woodall, W. H. (2009). Risk-adjusted monitoring of survival times. *Statistics in Medicine*, 28, 1386–1401.

- Shroyer, A. L. W., Coombs, L. P., Peterson, E. D., Eiken, M. C., Delong, E. R., Chen, A., Ferguson, T. B., Grover, F. L., & Edwards, F. H. (2003). The society of thoracic surgeons: 30-day operative mortality and morbidity risk models. *Annals Thoracic Surgery*, 75, 1856–1865.
- Sousa, P., Uva, A. S., & Pinto, F. (2008). Risk-adjustment model in health outcomes evaluation: A contribution to strengthen assessment towards quality improvement in interventional cardiology. *International Journal for Quality in Health Care*, 20(5), 324–330.
- Steiner, S. H., Cook, R. J., Farewell, V. T., & Treasure, T. (2000). Monitoring surgical performance using risk-adjusted cumulative sum charts. *Biostatistics*, 1(4), 441–452.
- Steiner, S. H., & Jones, M. (2010). Risk-adjusted survival time monitoring with an updated exponentially weighted moving average (EWMA) control chart. *Statistics in Medicine*, 29, 444–454.
- Tu, J. V., Jaglal, S. B., & Naylor, C. D. (1995). Multicenter validation of a risk index for mortality, intensive care unit stay, and overall hospital length of stay after cardiac surgery. *Circulation*, 91, 677–684.
- Woodall, W. H. (2006). The use of control charts in health-care and public-health surveillance. *Journal of Quality Technology*, 38(2), 89–104.
- Zeng, L., & Zhou, S. (2011). A Bayesian approach to risk-adjusted outcome monitoring in healthcare. *Statistics in Medicine*, 30(29), 3431–3446.
- Zimmerman, D. R. (2003). Improving nursing home quality of care through outcomes data: The MDS quality indicators. *International Journal of Geriatric Psychiatry*, 18, 250–257.

Quality and Reliability Management and Its Applications

Pham, H. (Ed.)

2016, XVII, 451 p., Hardcover

ISBN: 978-1-4471-6776-1