Research Article

Statistics in Medicine

Received 6 October 2009, Accepted 26 July 2011 Published online 3 October 2011 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.4374

A Bayesian approach to risk-adjusted outcome monitoring in healthcare

L. Zeng[a*†] and S. Zhou[b]

Clinical outcomes are commonly monitored in healthcare practices to detect changes in care providers' performance. One key challenge in outcome monitoring is the need of adjustment for patient base-line risks. Various control charting methods have been developed to conduct risk-adjusted outcome monitoring, but they all rely on the availability of a large number of historical data. We propose a Bayesian approach to this type of monitoring for cases where historical data are not available. In our approach, detection of change is formulated as a model-selection problem and solved using a popular Bayesian tool for variable selection, the Bayes factor. Issues in decision-making about whether there is a change point in the observed patient outcomes are addressed, including specification of priors and computation of Bayes factors. This approach is applied to a real data set on cardiac surgeries, and its performance under different parameter scenarios is studied through simulations.

Keywords: Bayes factor; logistic regression model; Markov chain Monte Carlo (MCMC); outcome monitoring; risk adjustment

1. Introduction

As important indicators of quality of care, clinical outcomes are measured in all sorts of healthcare practices. Examples are patient mortality, longevity, complications of disease, and physical functional status. By monitoring such outcomes, changes in the performance of care providers can be detected promptly, which is a critical step in controlling quality of care.

One challenging aspect of outcome monitoring in the medical context is adjusting for differences in patient case mix, called *risk adjustment* [1, 2]. Unlike products in manufacturing processes, which are relatively homogeneous in nature, patients vary a lot in their characteristics, and thus bear different base-line risks to adverse outcomes. For example, sicker patients tend to experience worse outcomes, even with excellent care, than their healthier counterparts. The patient base-line risks must be taken into account to fairly assess the performance of care providers. Outcome monitoring with such an adjustment is called risk-adjusted monitoring.

This topic has received a lot of attention in recent years, and a variety of methods have been proposed to conduct this type of monitoring, the majority of which focus on surgical care such as cardiac surgeries. Figure 1 shows a typical data series collected in centers for cardiac surgery, where the outcomes are binary attributes indicating the survival/death (denoted as S/D in the figure) of patients within a certain period after their operations, and the patient base-line risk is measured by a Parsonnet score, which combines the effects of important patient characteristics, or risk factors, such as age, gender, and diabetic status. Such data are usually available for each single surgeon and the goal of outcome monitoring is to capture the changes in his/her performance. The popularity of surgeries in the research is due largely to the motivation of well publicized reports of cases where high rates of surgical complications remained undetected for an undue length of time [3], and the availability of a common recognition of patient

[a]*Department of Industrial and Manufacturing Systems Engineering, The University of Texas at Arlington, Arlington, TX 76019, USA*
[b]*Department of Industrial and Systems Engineering, The University of Wisconsin–Madison, Madison, WI 53706, USA*
**Correspondence to: Li Zeng, Department of Industrial and Manufacturing Systems Engineering, The University of Texas at Arlington, 500 West First Street, PO Box 19017, Arlington, TX 76019, USA.*
[†]*E-mail: lzeng@uta.edu*

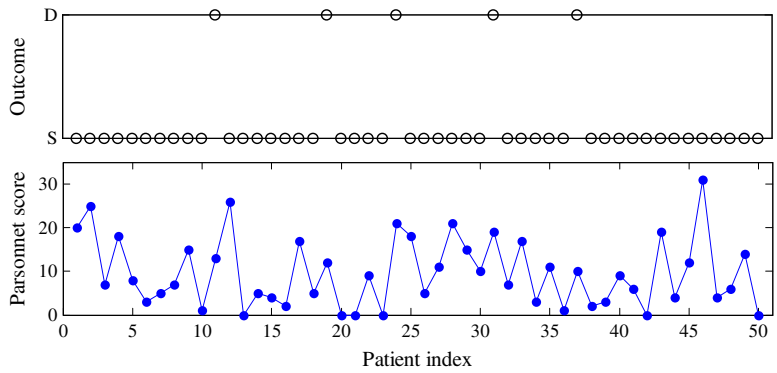


Figure 1. The data on cardiac surgeries: outcomes (upper) and risk scores (lower).

base-line risk factors to the outcomes. The proposed techniques include various risk-adjusted control charts, for example, cumulative risk-adjusted mortality charts [4], cumulative sum charts [3], and sets method/grass plots [5], which adjust the statistics in conventional outcome monitoring by the Parsonnet scores. Grigg and Farewell [6] gave an excellent review of them, and Woodall [7] commented on the issues with them.

These techniques are essentially extensions of statistical process control (SPC) charts that have been widely used to monitor industrial processes. A fundamental assumption of SPC charts is that there is a benchmark performance of the process, called normal/in-control performance, and a change means a deviation from it. Consequently, a monitoring scheme typically consists of two phases of analysis: Phase I or retrospective analysis, where the in-control performance is established using historical data, and Phase II or prospective analysis, where subsequent new measurements are examined to detect deviations from the in-control performance.

To establish the in-control performance, however, a long historical series is needed, which may not be generally available in practice, considering the limited capacity of a care provider and the cost of patient data collection. Moreover, even if historical data are available, there is often no guarantee that they were produced in a stable time period in which performance of the process was maintained at a constant level (i.e., the in-control level). Therefore, monitoring schemes that do not need historical data and start inspection from the first patients are practically desirable.

This article proposes such a scheme, a Bayesian approach, to risk-adjusted outcome monitoring. Bayesian approaches have been developed for general change detection in many other instances such as climate study [8], short-run production processes [9] and political science [10]. Certain good features of these methods, for example, not relying on asymptotics, applying for any sample size, and ease of use, make them a fitting tool for our problem.

Specifically, as each outcome is obtained, the proposed approach draws an inference about whether there is a *change point* among the available data rather than a deviation from a predetermined benchmark. This problem is formulated as a model selection problem in Bayesian sense and solved using a popular Bayesian tool for variable selection, the *Bayes factor*. A simple procedure is also developed for the computation of Bayes factors, which is, in general, a challenging task, based on direct MCMC outputs.

The remainder of this paper is organized as follows. Section 2 presents the mathematical formulation of the problem of risk-adjusted outcome monitoring. Technical details of the proposed approach are given in Section 3, including the decision rule and the procedure for calculating Bayes factors. This is followed by a case study in Section 4 where this approach is applied to the data set as shown in Figure 1. Section 5 presents the results of a numerical study, which is designed to demonstrate the performance of this approach under different parameter scenarios. Section 6 concludes and discusses open issues. For convenience, the proposed approach will be elaborated using the surgical example throughout the paper, though it is a generic solution for performance monitoring in healthcare applications.

2. Formulation of problem

Let Y_i be the outcome of patient i, which takes value 1 if he/she dies within a certain period after an operation, and 0 otherwise. The patient's base-line risk to death is represented by a risk score, x_i, such

as the Parsonnet score. A commonly used model to predict patient mortality is the logistic regression model

$$\log \frac{p_i}{1-p_i} = a + bx_i \tag{1}$$

where p_i is the mortality probability of patient i. Let $\boldsymbol{\beta} = [a\ b]'$ be the vector of model parameters, and $X_i = [1\ x_i]'$. An alternative expression of (1) is

$$P(Y_i = y|\boldsymbol{\beta}) = \frac{e^{X_i'\boldsymbol{\beta}\cdot y}}{1+e^{X_i'\boldsymbol{\beta}}} \tag{2}$$

where $y = 0$ or 1, and $P(\cdot)$ is the notation for probability. Later $P(Y_i = y|\boldsymbol{\beta})$ will be written as $P(y|\boldsymbol{\beta})$ for short. The parameter $\boldsymbol{\beta}$ essentially represents the care provider's performance, and correspondingly, the goal of outcome monitoring is to detect the change in $\boldsymbol{\beta}$. Note that here risk adjustment is realized through monitoring the model parameters, which do not depend on patient base-line risks.

Before going to the formal definition of the problem, the scope of this study should be made clear: (i) *Prospective* monitoring is considered. This means that the length of data series being monitored is not fixed, but increases over time without limit. Whenever a new observation becomes available, inference will be drawn based on all available data about whether a change has occurred. (ii) *Abrupt/step* changes, which occur instantaneously and sustain afterwards, are of interest. Moreover, other than limiting detection to changes in the mean, a, of the logistic model as those risk-adjusted control charts [3, 4] do, changes in both a and/or b are within our consideration.

Suppose m observations, $\mathbf{y}_{[m]} = [y_1, y_2, \ldots, y_m]$, are available at the present moment. The outcomes follow the in-control parameter $\boldsymbol{\beta}_0$ up to a change point K, and then switch to a different parameter $\boldsymbol{\beta}_1$, that is,

$$\begin{aligned} Y_i &\sim LG(y|\boldsymbol{\beta}_0), \text{ for } i = 1, 2, \ldots, K; \\ Y_i &\sim LG(y|\boldsymbol{\beta}_1), \text{ for } i = K+1, K+2, \ldots, m. \end{aligned}$$

where $LG(y|\boldsymbol{\beta})$ denotes the density of the logistic regression model, and the data within each group are independent of one another. Here, the *a priori* unknown K can be any value in $\{1, 2, \ldots, m-1\}$. Consequently, the change detection problem can be formulated as a selection between the following two models

$$\begin{aligned} &M_0 : P(\mathbf{y}_{[m]}|\boldsymbol{\theta}_0), \boldsymbol{\theta}_0 = \{\boldsymbol{\beta}_0 : \boldsymbol{\beta}_0 \in R^2\} \\ &M_1 : P(\mathbf{y}_{[m]}|\boldsymbol{\theta}_1), \boldsymbol{\theta}_1 = \{(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, K) : \boldsymbol{\beta}_0 \in R^2, \boldsymbol{\beta}_1 \in R^2, 1 \leqslant K \leqslant m-1\} \end{aligned} \tag{3}$$

where R is the real number field, and $P(\mathbf{y}_{[m]}|\boldsymbol{\theta}_t), t = 0, 1$, is the probability density of $\mathbf{y}_{[m]}$ under M_t. Given the independence of data, the expressions of the densities can be easily obtained

$$P(\mathbf{y}_{[m]}|\boldsymbol{\theta}_0) = \prod_{i=1}^{m} P(y_i|\boldsymbol{\beta}_0) = \prod_{i=1}^{m} \frac{e^{X_i'\boldsymbol{\beta}_0\cdot y_i}}{1+e^{X_i'\boldsymbol{\beta}_0}} \tag{4}$$

$$P(\mathbf{y}_{[m]}|\boldsymbol{\theta}_1) = \prod_{i=1}^{K} P(y_i|\boldsymbol{\beta}_0) \prod_{i=K+1}^{m} P(y_i|\boldsymbol{\beta}_1) = \prod_{i=1}^{K} \frac{e^{X_i'\boldsymbol{\beta}_0\cdot y_i}}{1+e^{X_i'\boldsymbol{\beta}_0}} \prod_{i=K+1}^{m} \frac{e^{X_i'\boldsymbol{\beta}_1\cdot y_i}}{1+e^{X_i'\boldsymbol{\beta}_1}} \tag{5}$$

Accordingly, the monitoring consists of two tasks: (i) solving the model selection problem in (3), that is, determining the existence of the change point K and (ii) estimating K if M_1 is selected.

3. The Bayesian approach

3.1. The Bayes factor

The model selection problem in (3) can be solved using the *Bayes factor*, which is a popular Bayesian tool for variable selection. The Bayes factor, denoted by BF_m, of competing models M_0 and M_1 is defined as the ratio of their corresponding marginal likelihoods [11]

$$BF_m = \frac{P(\mathbf{y}_{[m]}|M_1)}{P(\mathbf{y}_{[m]}|M_0)} = \frac{\int P(\boldsymbol{\theta}_1|M_1)P(\mathbf{y}_{[m]}|\boldsymbol{\theta}_1, M_1)d\boldsymbol{\theta}_1}{\int P(\boldsymbol{\theta}_0|M_0)P(\mathbf{y}_{[m]}|\boldsymbol{\theta}_0, M_0)d\boldsymbol{\theta}_0} \tag{6}$$

where $P(\boldsymbol{\theta}_t|M_t), t = 0, 1$, is the prior distribution under model M_t. Essentially, the Bayes factor is the Bayesian version of the likelihood ratio, except that the marginal likelihood is obtained by integrating (not maximizing) over the parameter space. Interested readers are referred to [12] for a comprehensive review of the use of this tool in various applications.

The Bayes factor is a summary of the evidence provided by the data in favor of one model as opposed to another. It treats the two competing hypotheses (models) equally, rather than giving one preferred status (the 'null hypothesis') as frequentist methods do [12]. Jeffreys [13] suggested a set of cutoff values for the Bayes factor as follows:

BF_m	Evidence against M_0
< 1	Negative
$1 \sim 3$	Barely worth mentioning
$3 \sim 10$	Substantial
$10 \sim 30$	Strong
$30 \sim 100$	Very strong
> 100	Decisive

As indicated in (6), there are two issues involved in the use of Bayes factors: specification of priors and computation of Bayes factors, which will be addressed next.

3.2. *Specification of priors*

To calculate the Bayes factor, the priors under the two hypothetical models, that is, $P(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, K|M_1)$ and $P(\boldsymbol{\beta}_0|M_0)$, need to be specified. Three assumptions can be made to simplify this problem: first, assume that under M_0,

$$P(\boldsymbol{\beta}_0|M_0) = \pi(\boldsymbol{\beta}) \tag{7}$$

It is also reasonable to assume that under M_1, K is independent of $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_1$,

$$P(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, K|M_1) = P(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1|M_1) \cdot \pi(K|M_1) \tag{8}$$

Furthermore, we can assume that under $M_1, \boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_1$ independently and identically follow the prior under M_0, that is,

$$P(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1|\boldsymbol{\beta}, M_1) = P(\boldsymbol{\beta}_0|\boldsymbol{\beta}, M_1) \cdot P(\boldsymbol{\beta}_1|\boldsymbol{\beta}, M_1) \tag{9}$$

$$P(\boldsymbol{\beta}_0|\boldsymbol{\beta}, M_1) = \pi(\boldsymbol{\beta}) \tag{10}$$

$$P(\boldsymbol{\beta}_1|\boldsymbol{\beta}, M_1) = \pi(\boldsymbol{\beta}) \tag{11}$$

As a result, the specification of priors boils down to specifying $\pi(K|M_1)$ and $\pi(\boldsymbol{\beta})$.

For K, a simple and common choice is

$$K|M_1 \sim (\text{discrete})\text{uniform}\{1, 2, \ldots, m-1\}$$

as there is generally no information about the location of the change point.

The specification of priors for the parameter of the logistic regression model, $\boldsymbol{\beta}$, has been considered in some studies. The choice depends on the availability of prior information such as historical data and expert knowledge. For example, Chen *et al*. [14] proposed a method for specifying the prior when closely related historical data are available, while Chaloner *et al*. [15] and Kadane and Wolfson [16] demonstrated how to elicit priors from expert opinions. When there is no historical data or expert assistance, normal prior [17, 18], conjugate prior [19], Jeffreys's prior [20], and flat prior [21, 22] have been applied. By integrating the existing work, we propose to use two convenient priors, the conditional means prior (CMP) for use when prior information is available and the truncated flat prior for use otherwise. Details of the priors can be found in Appendix A.

3.3. *Computation of Bayes factors*

The definition in (6) indicates that calculation of Bayes factors eventually boils down to computing the two marginal likelihoods as integrals over the parameter space. This is conventionally realized

using numerical integration techniques and ordinary Monte Carlo methods [12], which are very time-consuming and difficult to implement in high-dimensional cases. MCMC methods have been considered as a faster and more powerful solution recently [23], but the existing algorithms are very complex because marginal likelihoods cannot be estimated by direct posterior sampling [24]. This is because the marginal likelihood is the normalizing factor of the posterior density, rather than a posterior probability, which cannot be calculated using posterior samples. In this study, taking advantage of the characteristics of the problem, we developed a procedure through which Bayes factors can be obtained easily and directly from posterior sampling. Moreover, the change point can also be estimated simultaneously.

This procedure builds upon a simple idea that the hypothesis of no change, that is, M_0 in (3), can also be represented as '$K = m$', and in this way the model selection problem in (3) can be reformulated under a unified parameter setting, that is,

$$\begin{aligned} M_0 &: P(\mathbf{y}_{[m]}|\boldsymbol{\theta}_0), \boldsymbol{\theta}_0 = \{(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, K) : \boldsymbol{\beta}_0 \in R^2, \boldsymbol{\beta}_1 \in R^2, K = m\} \\ M_1 &: P(\mathbf{y}_{[m]}|\boldsymbol{\theta}_1), \boldsymbol{\theta}_1 = \{(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, K) : \boldsymbol{\beta}_0 \in R^2, \boldsymbol{\beta}_1 \in R^2, 1 \leqslant K \leqslant m-1\} \end{aligned} \tag{12}$$

Correspondingly, the unified prior of K is

$$K \sim (\text{discrete}) \text{ uniform}\{1, 2, \ldots, m-1, m\} \tag{13}$$

The advantage is that under the new formulation, the Bayes factor has a much simpler expression, as given in the following theorem:

Theorem
For the model selection problem in (12) and under the priors in (7)–(11) and (13), the Bayes factor can be simplified to

$$BF_m = \frac{P(K < m|\mathrm{y}_{[m]})/(m-1)}{P(K = m|\mathrm{y}_{[m]})} \tag{14}$$

where $P(\cdot|\mathrm{y}_{[m]})$ is the posterior probability.

Proof of this theorem can be found in Appendix B. The simple formula in (14) is very intuitive: as implied by (12), the selection between M_1 and M_0 essentially concerns whether the change point is smaller than or equal to m. Thus, it is no surprise that the Bayes factor weighs the posterior probabilities of $K < m$ and $K = m$. Because $K < m$ contains $m-1$ possible locations (i.e., $K = 1, 2, \ldots, m-1$), the Bayes factor, more precisely, weighs the average probability of all possible locations of the change point and the probability that there is no change point. In another perspective, (14) also suggests that inference based on the Bayes factor is essentially based on the posterior distribution of K. A note about (14) is that it only holds when the change point is discrete as considered in this study. In cases where continuous time series are monitored (e.g., [8,9]), computation of Bayes factors has to follow the original definition in (6).

Based on (14), the Bayes factor can be calculated through the following procedure:

Procedure
I. Generate b samples $\{(\boldsymbol{\beta}_0^{(j)}, \boldsymbol{\beta}_1^{(j)}, K^{(j)}) : j = 1, \ldots, b\}$ from the joint posterior distribution

$$P(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, K|\mathrm{y}_{[m]}) \propto \pi(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, K) P(\mathrm{y}_{[m]}|\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, K) \tag{15}$$

where the prior $\pi(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, K) = \pi(\boldsymbol{\beta}_0) \cdot \pi(\boldsymbol{\beta}_1) \cdot \pi(K)$, $\pi(K)$ follows (13), and (π $\boldsymbol{\beta}_0$) and (π $\boldsymbol{\beta}_1$) follow (10) and (11). The density function in (15) is

$$\begin{aligned} P(\mathbf{y}_{[m]}|\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, K) &= \left[\prod_{i=1}^{K} P(y_i|\boldsymbol{\beta}_0) \prod_{i=K+1}^{m} P(y_i|\boldsymbol{\beta}_1)\right] \cdot 1_{\{K<m\}} + \left[\prod_{i=1}^{m} P(y_i|\boldsymbol{\beta}_0)\right] \cdot 1_{\{K=m\}} \\ &= \left[\prod_{i=1}^{K} \frac{e^{X_i'\boldsymbol{\beta}_0 \cdot y_i}}{1 + e^{X_i'\boldsymbol{\beta}_0}} \prod_{i=K+1}^{m} \frac{e^{X_i'\boldsymbol{\beta}_1 \cdot y_i}}{1 + e^{X_i'\boldsymbol{\beta}_1}}\right] \cdot 1_{\{K<m\}} + \left[\prod_{i=1}^{m} \frac{e^{X_i'\boldsymbol{\beta}_0 \cdot y_i}}{1 + e^{X_i'\boldsymbol{\beta}_0}}\right] \cdot 1_{\{K=m\}} \end{aligned}$$

where $1_{\{B\}}$ is an indicator function that is equal to 1 if B is true.

II. Count the number of value $i, i = 1, \ldots, m$, in the sequence $\{K^{(j)} : j = 1, \ldots, b\}$. Denote this number as $N(i)$. Accordingly, BF_m can be estimated by

$$\hat{BF}_m = \frac{b - N(m)}{(m-1) \cdot N(m)} \tag{16}$$

And the point estimator of the change point is

$$\hat{K} = \underset{1 \leqslant i \leqslant m}{\arg \max} [N(i)] \tag{17}$$

Proof of (16) and (17) is given in Appendix C. This procedure bears three advantageous features: (i) it transforms the Bayes factor from a ratio of marginal likelihoods (i.e., $P(\mathbf{y}_{[m]}|\cdot)$) to that of posterior probabilities (i.e., $P(\cdot|\mathbf{y}_{[m]})$), which can be easily calculated through direct posterior sampling. (ii) The unified parameter setting makes it possible to calculate the two marginal likelihoods in (6) simultaneously using posterior samples of the common parameters, rather than calculate them separately as conventional methods do. This not only simplifies the computation, but also improves its accuracy as the error in the denominator and that in the numerator are from the same set of samples and thus counteract each other. (iii) The two tasks involved in solving the model selection problem in (12), that is, testing whether there is a change point and estimating the location of the change point, are carried out simultaneously using posterior samples.

3.4. *Implementation of the proposed approach*

The steps to implement the proposed approach in practice are summarized as follows:

Step 1: Specify priors for the change point K and the parameter of the logistic regression model $\boldsymbol{\beta}$.

- The prior of K follows (13).
- When prior information, related historical data or domain expert knowledge, is available, a conditional means prior of $\boldsymbol{\beta}$ will be specified following the procedure described in Appendix A.1. When there is no prior information, a truncated flat prior given in Appendix A.2 will be specified by common sense.

Step 2: Conduct change detection following the procedure illustrated in Figure 2: as each new data point (x_m, y_m), $m \geqslant 2$, is obtained, BF_m is calculated through the procedure presented in Section 3.3 and compared with a preset threshold η. If $BF_m > \eta$, M_1 will be selected, indicating that a change, either a performance improvement or deterioration, has occurred, and then the change point will be estimated. Otherwise, M_0 will be selected, meaning that there has been no change till now.

4. Case study

The proposed approach has been applied to a set of data from a UK center for cardiac surgery, part of which is shown in Figure 1. The whole data set contains information on each patient during 1992–1998, including time of the operation, surgeon performing the operation, type of surgical procedure, patient Parsonnet score, and 30-day mortality following the operation. The average patient risk score is about 30. These data have been used in several related studies (e.g., [3, 6]) where risk-adjusted control charts are developed for performance monitoring.

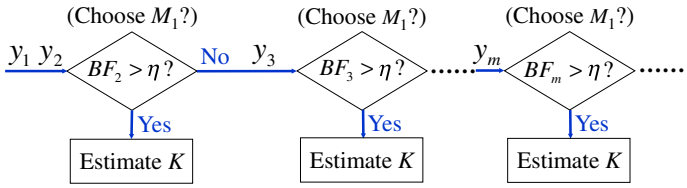


Figure 2. Change detection procedure based on Bayes factors.

4.1. *Specification of prior of* $\boldsymbol{\beta}$

Conditional means prior of $\boldsymbol{\beta}$ was specified in the study as prior information often exists in medical practices. Following the procedure in Appendix A.1, two typical risk levels were first selected: $x_{(1)} = 0$, $x_{(2)} = 30$, representing the healthiest patients and those with average risk to death, respectively. To find the hyper-parameters, we assumed that the mode and 95^{th} percentile are $p_m = 0.025$, $p_{95} = 0.1$ for $p_{(1)}$, and $p_m = 0.2$, $p_{95} = 0.6$ for $p_{(2)}$, which are consistent with the identified model parameters in the literature [3]. Because γ_t, $t = 1, 2$, can be expressed as a function of α_t by (A2), we obtained the 95^{th} percentile of a Beta distribution under a series of α_t values through simulation. Figure 3 shows the α values and the 95^{th} percentiles of corresponding Beta distributions. It is easy to find that when $\alpha_1 = 2.2$, $\alpha_2 = 1.9$ and thus $\gamma_1 = 47.8$ and $\gamma_2 = 5.8$, the corresponding 95^{th} percentiles are equal to the specifications, 0.1 and 0.6.

4.2. *Posterior sampling*

As shown in Section 3.3, the key step in calculating Bayes factors is sampling from the posterior distribution in (15), which can be realized via various MCMC methods such as Gibbs sampling and Metropolis–Hastings (MH) algorithm [11]. A recently emerging method, called *slice sampling* [25], is preferred here for its convenience in use. The slice sampler only needs two inputs to work: the posterior distribution to be sampled from and the initial value of the random sample sequence, which can be casually picked from the support of the priors. This is much simpler than using other MCMC methods, say, MH algorithm, where an appropriate proposal distribution needs to be selected with care. In addition, this method bears the major virtue of MH algorithm as well, that is, it is able to generate samples from a posterior distribution without knowing its normalization factor, which is often very difficult to compute. In other words, when using this method, we only need to provide the prior, that is, $\pi(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, K)$, and the density function, that is, $P(\mathbf{y}_{[m]}|\, \boldsymbol{\beta}_0, \boldsymbol{\beta}_1, K)$.

It has a minor limitation, however, that it only works on continuous distributions, and thus cannot be directly applied here for posterior sampling of the discrete K. As a simple way to conquer this issue, an ancillary variable λ is defined, which follows a continuous uniform distribution on (0, 1] such that $K = \lceil \lambda \cdot m \rceil$ follows a discrete uniform distribution on $\{1, \ldots, m\}$, where $\lceil c \rceil$ is the ceiling of c, that is, the smallest integer greater than or equal to c. Consequently, the posterior samples of K can be obtained indirectly through sampling from $P(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \lambda|\mathbf{y}_{[m]})$, which can be realized via slice sampler. In our study, this sampling was conducted using the *slicesample* function in MATLAB, and the Bayes factor was calculated based on 500,000 posterior samples with 10,000 burn-ins.

4.3. *Results*

A data series from a single surgeon, as shown in Figure 4, were used in the monitoring. The y axis in the figure represents the Pasonnet scores of patients, and the red dots indicate deaths. The inspection started from the second patient and yielded a sequence of Bayes factors as shown in Figure 5(a). Given $\eta = 7$, a change was detected at the 654^{th} observation. Figure 5(b) shows the sample posterior distribution of K based on all the 654 observations that have been inspected. The mode of this distribution, which is the

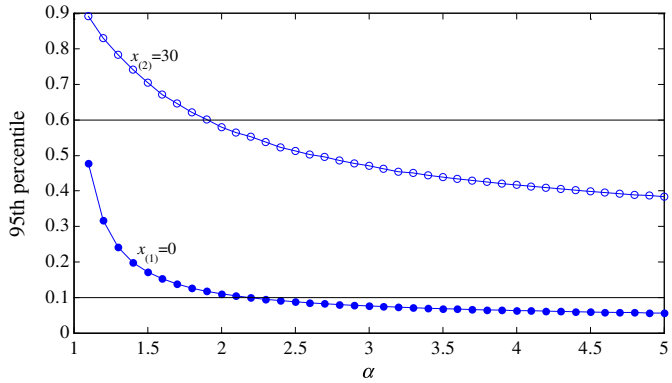


Figure 3. α values and corresponding 95^{th} percentiles.

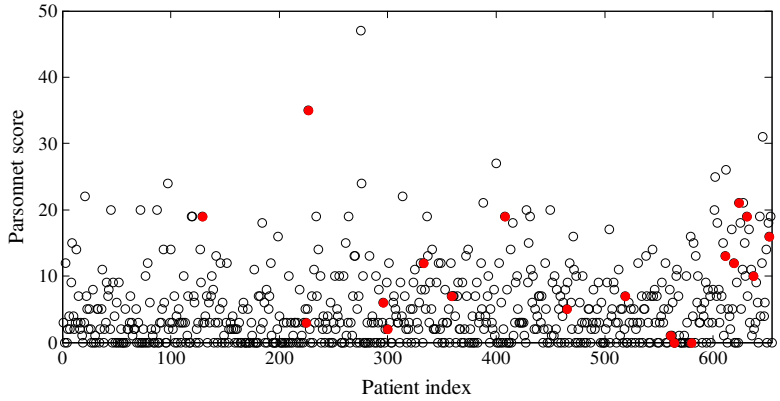
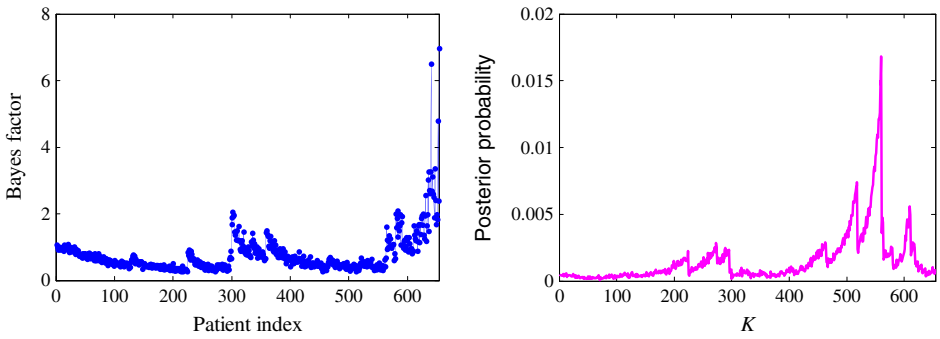


Figure 4. The data monitored in the case study.



(*a*) Bayes factors (*b*) Posterior of K

Figure 5. The resulting Bayes factors and sample posterior distribution of K.

estimator of K, is 560. Returning to Figure 4, it is clear that the surgeon's performance seems to have changed around the 560^{th} patient.

In addition, while the proposed approach is designed for prospective change detection, it can also be used for retrospective change detection, which is based on a fixed data set. This is demonstrated using the data series in the upper panel of Figure 6, which consists of patient outcomes from two surgeons who exhibited apparently different performances, with the first 600 points from surgeon1 and the following 510 points from surgeon2. Applying the procedure in Section 3.3 to these data yielded a $BF = 21.5$,

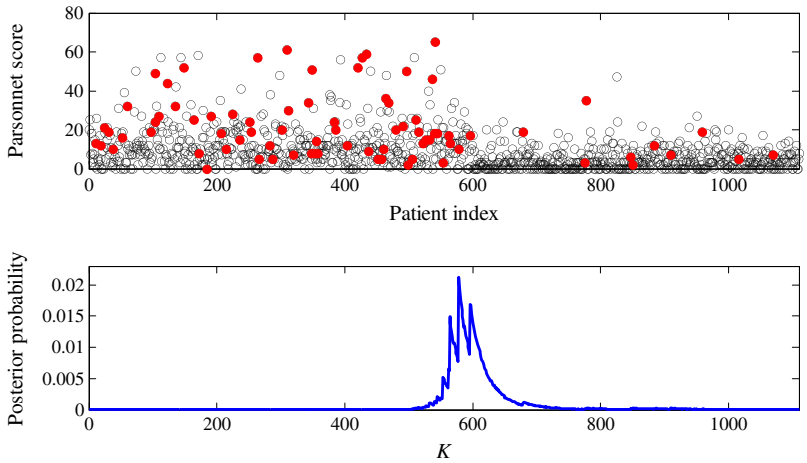


Figure 6. Retrospective change detection: data (upper) and resulting posterior of K (lower).

and a sample posterior distribution of K shown in the lower panel of Figure 6. The estimated change point is 578, which is very close to the truth.

5. Numerical study

A numerical study has been conducted to provide readers more insights on how the proposed approach performs under different parameter scenarios. Specifically, two problems have been studied: the effects of important factors affecting the performance of this approach and its performance under typical scenarios. In this section, the basic setting used in the study will first be introduced, and then simulation results will be summarized.

5.1. *Basic setting*

In the simulations, the true base-line parameter $\boldsymbol{\beta}_0 = [a_0, b_0]'$ was assumed to be $a_0 = -5$, $b_0 = 0.3$, and the patient risk scores ranged from 0 to 20. Under this setting, the lowest, average, and highest risks of death are 0.67%, 11.92%, and 73.1%, respectively, which are similar to the reported cases of cardiac surgeries [3] and open heart surgeries [26].

To specify the prior of $\boldsymbol{\beta}$, two risk levels were selected: $x_{(1)} = 0$, $x_{(2)} = 10$, corresponding to the lowest and average risks respectively. By varying the expert opinions within reasonable ranges, different CMPs were obtained following the procedure in Appendix A.1 and used in the simulations. It turned out that the performance of this approach is not considerably sensitive to the priors. The typical results under these CMPs will be presented next.

5.2. *Effects of important affecting factors*

In general, the performance of a change detection scheme is affected by two sets of factors: characteristics (i.e., magnitude and starting location) of the change and available information (i.e., the number of observations). Effects of these factors are investigated in this study.

From Section 3.3, we have learned that the posterior distribution of K can be used as an evidence of change. This provides a convenient way to demonstrate the effects of affecting factors. Three cases were created by varying the magnitude and starting location of the change, as listed in Table I. In each case, a number of data series were generated, and for each series, posterior sampling was conducted when $m = 40, 60, 80$. The characteristic behaviors of the resulting sample posterior distributions of K are shown in Figure 7.

In each plot of Figure 7, the height of the bar corresponding to i, $1 \leqslant i \leqslant m$, denotes the posterior probability $P(K = i|\mathbf{y}_{[m]})$. The two red bars, one at the true change point (called Bar1) and the other at m (called Bar2), represent $P(K < m|\mathbf{y}_{[m]})/(m-1)$, that is, the average height of bars at 1, 2,...,$m-1$, and $P(K = m|\mathbf{y}_{[m]})$, respectively. According to (14), the relative height of these two bars represents the value of BF_m. The information in Figure 7 can be summarized as follows:

(i) *The evidence of change becomes stronger as more data become available.* In each case, as m increases, the relative height of Bar1 to Bar2 becomes higher, or equivalently, the Bayes factor becomes larger. Another manifestation of the stronger evidence is the fact that some bars become more salient. The greater concentration of K on the area around the true change point (most apparently seen in Case 1) also suggests that with more data, inference of the change point will become more accurate. These are consistent with our expectation. Essentially, Bayesian methods are characterized by their ability in approaching truth through fully utilizing data, and better inference will be achieved as more data become available.

(ii) *The approach performs better in detecting larger changes.* Comparing Case 1 and Case 2, under a larger change (i.e., Case 1), the evidence of change is stronger and the inference on the change

Table I. Specification of parameters in the three cases.

	Parameters	Interpretation
CASE 1	$\boldsymbol{\beta}_1 = [-2.5, 0.3]$, $K = 30$	large magnitude, late location
CASE 2	$\boldsymbol{\beta}_1 = [-4, 0.3]$, $K = 30$	small magnitude, late location
CASE 3	$\boldsymbol{\beta}_1 = [-2.5, 0.3]$, $K = 30$	large magnitude, early location

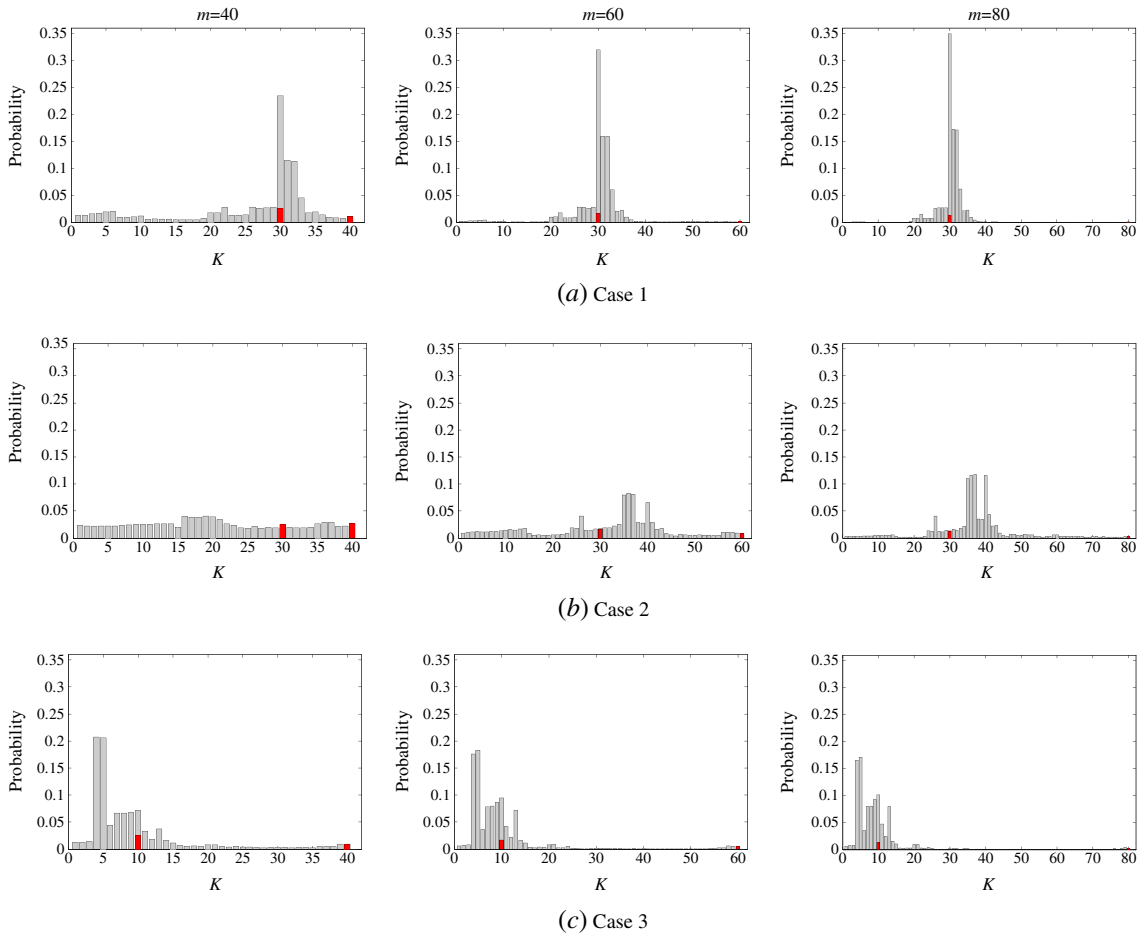


Figure 7. Characteristic sample posterior distributions of K.

point is more accurate with the same amount of data. This seems also natural because larger changes tend to manifest themselves easier.

(iii) *The approach does not perform well in detecting changes occurring at an early time.* Comparing Case 1 and Case 3, under an early change (i.e., Case 3), the evidence of change is clear, but inference on the change point is misleading. This is due to the lack of data from the prechange model.

5.3. Performance under different scenarios

To measure the performance of the proposed approach quantitatively, a set of parameter scenarios have been created, as given in Table II. There were a total of 16 cases, with one no-change case and 15 cases with change. Among the latter, three starting locations and five magnitudes of the change are considered. The performance of the approach under each scenario is measured based on 200 replications of simulation.

In each replication, the procedure described in Section 3.4 was applied to a randomly generated data series under this scenario, and 8 values of η, from 3 to 10, were used to make a decision on the resulting Bayes factors. In the no-change case, the monitoring started from $m = 2$, whereas in cases with change, it started right after the change occurred. When a change was detected, the present index of observation was recorded as the *signal location*. Because computation of Bayes factors is time intensive, the monitoring was stopped as m reached 300 rather than keeping it running until a signal is obtained. In other words, the simulation stopped when a change was detected or when 300 observations had been inspected. The time needed for calculating the Bayes factor is linearly dependent on the number of available observations. For example, it takes about 1.5 min when $m = 300$. This speed is not very fast, but still acceptable for practical use considering the slow accumulation of patient outcomes in healthcare practices.

Table II. Setting of parameters in the simulation.

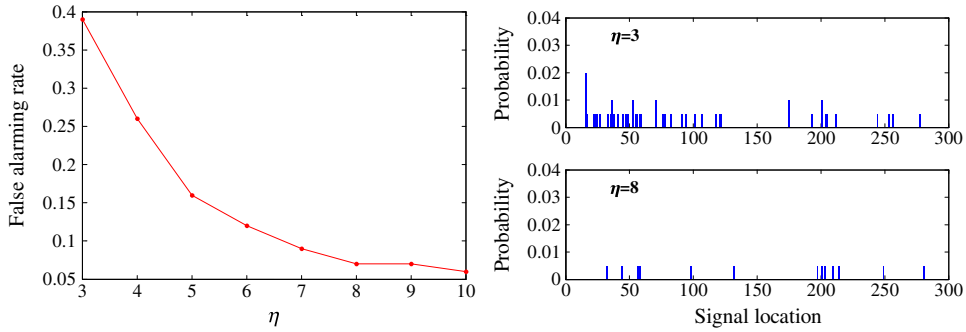
Fix parameters
(1) normal model parameters: $a_0 = -5, b_0 = 0.3$
(2) risk scores: $x \sim$uniform[0, 20], $\mu = 10$
Change scenarios
(1) location of change point : $K = 30, 50, 70$
(2) magnitude of change:
$M_1 : a_1 = -4.3, b_1 = 0.3$
$M_2 : a_1 = -3.6, b_1 = 0.3$
$M_3 : a_1 = -3, b_1 = 0.3$
$M_4 : a_1 = -5, b_1 = 0.4$
$M_5 : a_1 = -4.3, b_1 = 0.4$
Decision threshold
(1) $\eta = 3, 4, 5, 6, 7, 8, 9, 10$

The following measures were used to evaluate the performance under each scenario:

Signal rate: the percentage of replications (among the 200 replications) in which a change is detected. This represents the *false alarm rate* in the no-change case, and the (detection) *power* in cases with change.

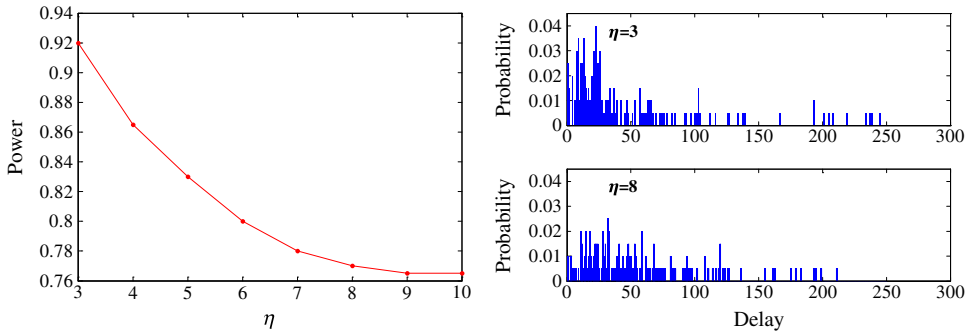
Distribution of signal locations: this is the distribution of locations of false alarms in the no-change case, and the distribution of delays in capturing the change in cases with change.

Here we use the results in two representative cases, the no-change case and the case under $K = 50$ and M_2 (simply called 'the change case' hereafter), to summarize our findings. Figure 8 displays the results in the no-change case, including the false alarm rate with respect to different values of η and the distribution of signal locations under $\eta = 3$ and 8. Figure 9 shows the power and distribution of delays in the change case.



(*a*) False alarm rate (*b*) Distribution of signal locations

Figure 8. Performance of the proposed approach in the no-change case.



(*a*) Power (*b*) Distribution of delays

Figure 9. Performance of the proposed approach in the change case.

From Figures 8(a) and 9(a), we can see that the higher the threshold η is, the smaller the false alarm rate, and the lower the power. When the lowest threshold is used, the false alarm rate is as high as 0.39, while the power is 0.92 indicating an excellent performance. In contrast, when the highest threshold is used, the false alarm rate reduces to as low as 0.05, but the power decreases as well to 0.76. Overall, the performance of the proposed procedure is reasonable, considering the fact that our analysis is not based on a large historical data set, binary outcomes are used to make inference rather than more informative continuous data as used in other applications (e.g., [8,9]), and patient base-line risks vary a lot from one to another.

According to Figure 8(b), in the no-change case, false alarms largely occur at the beginning of the monitoring. This is easy to understand as inference is closer to the truth when more data are available. In the change case, according to Figure 9(b), most signals appear within the area very close to the true change point, meaning that the change was detected promptly. For example, even under a very high threshold (i.e., $\eta = 8$), 60%, 75%, and 85% signals occur within 50, 80, and 100 patients after the change takes place, respectively.

Because the choice of η affects the performance, general guidelines are provided as follows: as the simulation results suggest, a higher threshold leads to a lower false alarm rate and a lower power, so the choice depends on the trade-off of the two types of errors in a specific application. For example, in cases where failing to detect a change can result in serious consequences, such as critical medical contexts like cardiac surgeries and intensive care unit, we can choose a relatively small threshold, for example, 4 or 5, to ensure a high detection power. In cases where false alarms are not tolerable, we have to pick a large threshold, for example, 7 or 8, at the cost of a considerable chance to miss changes of small magnitudes. In general, 6 is an appropriate choice as it can achieve acceptable performance in both aspects.

Results in other cases show similar phenomena, except for the differences because of the effects of occurring location and magnitude of change, as summarized in Section 5.2.

6. Conclusion and discussion

In this article, we proposed a Bayesian approach for risk-adjusted outcome monitoring in healthcare. Using this approach, our knowledge of a care provider's performance can be updated as each patient outcome becomes available, and thus changes in the performance can be detected promptly to avoid serious consequences. In the case study, the proposed approach is applied to a data set on cardiac surgeries. Results of a comprehensive numerical study are also presented to demonstrate the performance of this approach under different parameter scenarios and provide guidelines for its use in practice. This approach focuses on detecting the change of model parameters, and can be applied wherever a predictive model of the care provider's performance is available. Thus, while illustrated using binary outcomes in this article, it is essentially a generic solution for performance monitoring in various healthcare applications.

Several interesting open issues remain in this study. First, a way to enhance efficiency in change detection is to restrict the range of K under M_1 to a window $W = \{1 + \Delta_1, \ldots, m - \Delta_2\} \subset \{1, \ldots, m - 1\}$, where Δ_1 and Δ_2, $\Delta_1 \geqslant 1$, $\Delta_2 \geqslant 2$, are small integers, for example, 5, because it is not very likely to capture a change at the very beginning or end of the data. Fortunately, the results in this paper can be easily extended to such a restricted setting. Second, the Bayes factor used here is more precisely an overall Bayes factor that takes all available data into account. This might not be the best choice for prompt detection of changes as a change needs to take some sampling intervals to become apparent in an overall measure. A local Bayes factor can be defined similarly and utilized to reduce the delay in monitoring. In fact, the ratio between adjacent Bayes factors can be viewed as one type of such local measures as it represents the effect of the present observation. West [27] proposed the idea of cumulative Bayes factors (CBFs), which is a product of consecutive local Bayes factors, for process monitoring. The potential of CBFs has been proved by our simulations. However, to implement this idea, two critical parameters need to be determined: the appropriate number of local Bayes factors involved in a CBF and the threshold of CBFs used in decision making in the monitoring. These issues will be considered in a future study. Third, the proposed approach does not differentiate the change in the mean of the logistic regression model, that is, parameter a, and that in the coefficient, that is, parameter b. Sometimes, a differentiation is useful or necessary, as the two types of changes may have different implications in a specific application. How to adjust the approach to detect each type of change is a topic worth pursuing. Finally, in the numerical study, we ignored the effect of patient risk scores on the

performance of this approach by simply assuming them to follow a uniform distribution. It is also interesting to investigate how the distribution of risk scores affects the performance of outcome monitoring in the future.

Appendix A: Priors of $\boldsymbol{\beta}$

A.1. Conditional means prior

According to Kadane and Wolfson [16], one agreement that has been reached on elicitation of priors is that experts should be asked to assess only observable quantities such as probabilities or quantiles of the predictive distribution. In our case, it is obvious that eliciting information about mortality probabilities (i.e., p_i in (1)) is much easier than eliciting information about parameters of the logistic regression model (i.e. $\boldsymbol{\beta}$ in (2)). The CMP proposed by Bedrick *et al*. [28] just rests on this idea, which first specifies priors for typical mortality probabilities and then induces the prior of $\boldsymbol{\beta}$ through transformation.

Specifically, for a logistic regression model with one covariate, two typical risk scores, $x_{(1)}$ and $x_{(2)}$, $x_{(1)} < x_{(2)}$, need to be selected, whose corresponding mortality probabilities are

$$p_{(1)} = \frac{e^{X'_{(1)}\boldsymbol{\beta}}}{1 + e^{X'_{(1)}\boldsymbol{\beta}}}, \quad p_{(2)} = \frac{e^{X'_{(2)}\boldsymbol{\beta}}}{1 + e^{X'_{(2)}\boldsymbol{\beta}}}$$

Here, $x_{(1)}$ and $x_{(2)}$ should be within the range of observed risk scores, but far enough apart so that $p_{(1)}$ and $p_{(2)}$ can be reasonably treated to be independent. In our problem, this assumption is violated for the fact that a higher risk score normally leads to a higher mortality probability, and thus $p_{(1)} < p_{(2)}$, that is, they bear a positive correlation. In terms of the parameter, this is equivalent to the constraint $b > 0$. To reflect this correlation, we need to incorporate this constraint in the prior setting, as suggested in [28].

A common choice for the prior of mortality probabilities is the Beta distribution

$$p_{(1)} \sim \text{Beta}\,(\alpha_1, \gamma_1), \quad p_{(2)} \sim \text{Beta}\,(\alpha_2, \gamma_2)$$

And thus

$$\pi(p_{(1)}, p_{(2)}) \propto \prod_{t=1}^{2} p_{(t)}^{\alpha_t - 1}(1 - p_{(t)})^{\gamma_t - 1}$$

Using the change-of-variables technique, the induced CMP of $\boldsymbol{\beta}$ is

$$\pi(\boldsymbol{\beta}) \propto \prod_{t=1}^{2} p_{(t)}^{\alpha_t - 1}(1 - p_{(t)})^{\gamma_t - 1}\dot{p}_{(t)} = \prod_{t=1}^{2}\left(\frac{e^{X'_{(t)}\boldsymbol{\beta}}}{1 + e^{X'_{(t)}\boldsymbol{\beta}}}\right)^{\alpha_t}\left(\frac{1}{1 + e^{X'_{(t)}\boldsymbol{\beta}}}\right)^{\gamma_t} \tag{A1}$$

where $\dot{p}_{(t)}$ is the first derivative of $p_{(t)}$, $t = 1, 2$, with respect to $\boldsymbol{\beta}$.

In summary, there are two steps to obtain the CMP:

(i) Select two typical risk scores, $x_{(1)}$ and $x_{(2)}$, which are amenable to the expert (physician).
(ii) Specify the hyperparameters, $\alpha_t, \gamma_t, t = 1, 2$. For a formal way to do this, we can refer to the literature on priors of binomial models. Chaloner and Duncan [29] proposed a posterior mode method based on the idea that the mode of the Beta distribution, that is, $(\alpha_t - 1)/(\alpha_t + \gamma_t - 2)$, is more attractive to elicit than a moment such as mean. Specifically, given N hypothetical patients with risk score $x_{(t)}$, the expert will be asked to assess the most likely number of deaths among them.

The values of α_t and γ_t can be obtained through repeating such an assessment for a number of different values of N. Gavasakar [30] points out the drawbacks of the posterior mode method and develops a modified procedure, which is more accurate and easier to incorporate historical data. However, the existing methods are very complex, which may hinder the acceptance of them by healthcare practitioners. A simpler solution is to derive the hyperparametersfrom two specified probabilities/quantiles of the Beta distribution. An appropriate choice is the mode and 95$^{\text{th}}$ (or 90$^{\text{th}}$) percentile for their intuitive implications. Accordingly, the expert will be asked two questions in the elicitation:

- *What is the most likely probability of death for a patient with risk score* $x_{(t)}$?
- *What is the upper limit of your assessment?*

The answers to these will be taken as the mode and 95$^{\text{th}}$ percentile of Beta (α_t, γ_t), denoted as p_m and p_{95}, respectively. For example, if the expert's answer is that 'I believe this probability is probably 20%, and it should not be beyond 50%', then $p_m = 0.2$, and $p_{95} = 0.5$. The values of α_t and γ_t will then be obtained by solving

$$\frac{\alpha_t - 1}{\alpha_t + \gamma_t - 2} = p_m \tag{A2}$$

$$F^{-1}(0.95|\alpha_t, \gamma_t) = p_{95} \tag{A3}$$

where $F(\cdot|\alpha_t, \gamma_t)$ is the cumulative distribution function of the Beta distribution. The satisfying solution can be found by plotting $F^{-1}(0.95|\alpha_t, \gamma_t)$ for possible values of α_t and γ_t satisfying (A2). Note that p_m and p_{95} can either reflect the expert's belief or be based on historical data from related studies.

A.2. Truncated flat prior

When there is no prior information, a flat or noninformative prior, that is, $\pi(\boldsymbol{\beta}) \propto 1$, $\boldsymbol{\beta} \in R^2$, can be used. A special consideration here is that the Bayes factor is only defined when the marginal density of data, that is, $P(\mathbf{y}_{[m]}|M_t)$, $t = 0, 1$, is proper, or equivalently, $\pi(\boldsymbol{\beta})$ is proper. In other words, we cannot let $\boldsymbol{\beta} \in R^2$, but need to restrict it on a truncated support. This can be conducted via the simple method described below.

Following the idea of the CMPs, we can choose two risk levels, $x'_{(1)} = 0$ and $x'_{(2)} > 0$, whose corresponding mortality probabilities, $p'_{(1)}$ and $p'_{(2)}$, satisfy

$$p_{1l} \leqslant p'_{(1)} = \frac{1}{1 + e^{-a}} \leqslant p_{1u}, \quad \text{and} \quad p'_{(1)} < p'_{(2)} = \frac{1}{1 + e^{-(a + b x'_{(2)})}} \leqslant p_{2u}$$

Because no other information is available, the limits of the supports, p_{1l}, p_{1u} and p_{2u}, can only be specified by common sense, and thus may bear a wide span. For example, we can let the mortality probability of patients with zero risk score to be within [1%, 50%]. Then the truncated support of $\boldsymbol{\beta}$ can be obtained:

$$\log \frac{p_{1l}}{1 - p_{1l}} \leqslant a \leqslant \log \frac{p_{1u}}{1 - p_{1u}}, \text{ and } 0 < b \leqslant \frac{\log \frac{p_{2u}}{1 - p_{2u}} - a}{x'_{(2)}}$$

Therefore, the truncated flat prior of $\boldsymbol{\beta}$ is

$$\pi(\boldsymbol{\beta}) = \pi(a)\pi(b|a) = \text{uniform}\left[\log \frac{p_{1l}}{1 - p_{1l}}, \log \frac{p_{1u}}{1 - p_{1u}}\right] \cdot \text{uniform}\left(0, \frac{\log \frac{p_{2u}}{1 - p_{2u}} - a}{x'_{(2)}}\right] \tag{A4}$$

Appendix B: Proof of Theorem in Section 3.3

We first calculate the marginal likelihood under M_1:

$$P(\mathbf{y}_{[m]}|M_1) = \int P(\boldsymbol{\theta}_1|M_1) P(\mathbf{y}_{[m]}|\boldsymbol{\theta}_1, M_1) \mathrm{d}\boldsymbol{\theta}_1$$

From Section 3.2, K is independent of $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_1$, and $K|M_1 \sim \text{uniform}\{1, 2, ..., m-1\}$. Therefore,

$$\begin{aligned}
P(\mathbf{y}_{[m]}|M_1) &= \frac{1}{m-1} \sum_K \iint P(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1|M_1) P(\mathbf{y}_{[m]}|\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, K, M_1) \mathrm{d}\boldsymbol{\beta}_0 \mathrm{d}\boldsymbol{\beta}_1 \\
&= \frac{1}{m-1} \sum_K \iint P(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \mathbf{y}_{[m]}|K, M_1) \mathrm{d}\boldsymbol{\beta}_0 \mathrm{d}\boldsymbol{\beta}_1 \\
&= \frac{1}{m-1} \sum_K P(\mathbf{y}_{[m]}|K, M_1) \\
&= \frac{1}{m-1} \sum_{i=1}^{m-1} P(\mathbf{y}_{[m]}|K = i)
\end{aligned} \tag{A5}$$

Let $\xi = \sum_{i=1}^{m-1} P(\mathbf{y}_{[m]}|K=i)$. By (13),

$$\frac{1}{m}\xi = \sum_{i=1}^{m-1}\frac{1}{m}P(\mathbf{y}_{[m]}|K=i) = \sum_{i=1}^{m-1}\pi(K=i)\cdot P(\mathbf{y}_{[m]}|K=i)$$

$$= \sum_{i=1}^{m-1} P(\mathbf{y}_{[m]}, K=i) = \sum_{i=1}^{m-1} P(\mathbf{y}_{[m]})P(K=i|\mathbf{y}_{[m]})$$

$$= P(\mathbf{y}_{[m]})\sum_{i=1}^{m-1} P(K=i|\mathbf{y}_{[m]}) = P(\mathbf{y}_{[m]})\cdot P(K<m|\mathbf{y}_{[m]})$$

Substituting ξ into (A5) gives

$$P(\mathbf{y}_{[m]}|M_1) = \frac{1}{m-1}\xi = \frac{m}{m-1}P(\mathbf{y}_{[m]})\cdot P(K<m|\mathbf{y}_{[m]}) \tag{A6}$$

Similarly, we can get the marginal likelihood under M_0:

$$\begin{aligned} P(\mathbf{y}_{[m]}|M_0) &= \int P(\boldsymbol{\theta}_0|M_0)P(\mathbf{y}_{[m]}|\boldsymbol{\theta}_0, M_0)\mathrm{d}\boldsymbol{\theta}_0 \\ &= \sum_{K=m}\iint P(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1|M_0)P(\mathbf{y}_{[m]}|\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, K, M_0)\mathrm{d}\boldsymbol{\beta}_0\mathrm{d}\boldsymbol{\beta}_1 \\ &= P(\mathbf{y}_{[m]}|K=m) \\ &= m\cdot\left[\frac{1}{m}P(\mathbf{y}_{[m]}|K=m)\right] \\ &= m\cdot\pi(K=m)P(\mathbf{y}_{[m]}|K=m) \\ &= mP(\mathbf{y}_{[m]})\cdot P(K=m|\mathbf{y}_{[m]}) \end{aligned} \tag{A7}$$

Substituting (A6) and (A7) to (6) yields (14).

Appendix C. Proof of (16) and (17) in Section 3.3

Because $P(K<m|\mathbf{y}_{[m]}) = 1 - P(K=m|\mathbf{y}_{[m]})$, to obtain BF_m in (14), we only need to calculate $P(K=m|\mathbf{y}_{[m]})$, which is defined as

$$P(K=m|\mathbf{y}_{[m]}) = \sum_K 1_{\{K=m\}}P(K|\mathbf{y}_{[m]})$$

Because

$$P(K|\mathbf{y}_{[m]}) = \iint P(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, K|\mathbf{y}_{[m]})\mathrm{d}\boldsymbol{\beta}_0\mathrm{d}\boldsymbol{\beta}_1$$

We can get

$$P(K=m|\mathbf{y}_{[m]}) = \sum_K\iint 1_{\{K=m\}}P(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, K|\mathbf{y}_{[m]})\mathrm{d}\boldsymbol{\beta}_0\mathrm{d}\boldsymbol{\beta}_1$$

$$= E_{P(\boldsymbol{\beta}_0,\boldsymbol{\beta}_1,K|y_{[m]})}[1_{\{K=m\}}]$$

Here, $E_{P(\cdot)}[g(\cdot)]$ is the expectation of function $g(\cdot)$ with respect to $P(\cdot)$. This can be estimated by

$$\hat{P}(K=m|\mathbf{y}_{[m]}) = \frac{1}{b}\sum_{j=1}^{b} 1_{\{K^{(j)}=m\}} = \frac{N(m)}{b} \tag{A8}$$

where $\{K^{(j)} : j = 1,\ldots,b\}$ is the posterior samples of K. From (14), (16) results. Similarly, the posterior mode of K, that is, $\arg\max_{1\le i\le m}[P(K|\mathbf{y}_{[m]})]$, can be estimated by (17).

Acknowledgements

The authors gratefully appreciate Dr. Landon H. Sego at Pacific Northwest National Laboratory, Richland, WA, Professor Stefan H. Steiner at University of Waterloo, Waterloo, ON, Canada, and Professor William H. Woodall at Virginia Polytechnic Institute and State University, Blacksburg, VA, for the access to the cardiac surgery data. We also sincerely thank the associate editor and two anonymous reviewers for their constructive comments and suggestions.

References

1. Iezzoni LI (ed.). *Risk Adjustment for Measuring Healthcare Outcomes*, 3rd ed. Healthcare Administration Press: Chicago, IL, 1997.
2. Woodall WH, Grigg OA, Burkom HS. Research Issues and Ideas on Health-Related Surveillance. In *Frontiers in Statistical Quality Control 9*, Lenz HJ, Wilrich PT (eds). Physica-Verlag: Heidelberg, Germany, 2010; 145–155.
3. Steiner SH, Cook RJ, Farewell VT, Treasure T. Monitoring Surgical Performance Using Risk-adjusted Cumulative Sum Charts. *Biostatitics* 2000; **1**:441–452.
4. 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.
5. Grigg OA, Farewell VT. A Risk-Adjusted Sets Method for Monitoring Adverse Medical Outcomes. *Statistics in Medicine* 2004; **23**:1593–1602.
6. Grigg OA, Farewell VT. An Overview of Risk-Adjusted Charts. *Journal of the Royal Statistical Society Series A* 2004; **167**:523–539.
7. Woodall WH. The Use of Control Charts in Health-Care and Public-Health Surveillance. *Journal of Quality Technology* 2006; **38**:89–104.
8. Lee TCK, Zwiers FW, Hegerl GC, Zhang X, Tsao M. A Bayesian Climate Change Detection and Attribution Assessment. *Journal of Climate* 2005; **18**:2429–2440.
9. Tsiamyrtzis P, Hawkins DM. A Bayesian Scheme to Detect Changes in the Mean of A Short-Run Process. *Technometrics* 2005; **47**:446–456.
10. Spirling A. Bayesian Approaches for Limited Dependent Variable Change Point Problems. *Political Analysis* 2007; **15**:387–405.
11. Gelman A, Carlin JB, Stern HS, Rubin DB. *Bayesian Data Analysis*, 2nd ed. Chapman & Hall/CRC: Boca Raton, Florida, 2004.
12. Kass RE, Raftery AE. Bayes Factors. *Journal of the American Statistical Association* 1995; **90**:773–795.
13. Jeffreys H. *The Theory of Probability*, 2nd ed. Oxford University Press: Oxford, UK, 1961.
14. Chen M, Ibrahim JG, Yiannoutsos C. Prior Elicitation, Variable Selection and Bayesian Computation for Logistic Regression Models. *Journal of the Royal Statistical Society, Series B* 1999; **61**:223–242.
15. Chaloner K, Church T, Louis TA, Matts JP. Graphical Elicitation of A Prior Distribution for A Clinical Trial. *The Statistician* 1993; **42**:341–353.
16. Kadane JB, Wolfson LJ. Experiences in Elicitation. *The Statistician* 1998; **47**:3–19.
17. Gelfand AE, Sahu SK, Carlin BP. Efficient Parametrizations for Generalized Linear Mixed Models. In *Bayesian Statistics 5*. Oxford University Press: Oxford, 1996; 165–180.
18. Raftery AE. Approximate Bayes Factors and Accounting for Model Uncertainty in Generalized Linear Models. *Biometrika* 1996; **83**:251–266.
19. Chen M, Ibrahim JG. Conjugate Priors for Generalized Linear Models. *Statistica Sinica* 2003; **13**:461–476.
20. Chen M, Ibrahim JG, Kim S. Properties and Implementation of Jeffreys's Prior in Binomial Regression Models. *Journal of the American Statistical Association* 2008; **103**:1659–1664.
21. Zeger SL, Karim MR. Generalized Linear Models with Random Effects: A Gibbs Sampling Approach. *Journal of the American Statistical Association* 1991; **86**:79–86.
22. Albert JH, Chib S. Bayesian Analysis of Binary and Polychotomous Response Data. *Journal of the American Statistical Association* 1993; **88**:669–679.
23. Han C, Carlin BP. Markov Chain Monte Carlo Methods for Computing Bayes Factors: A Comprehensive Review. *Journal of the American Statistical Association* 2001; **96**:1122–1132.
24. Chib S. Marginal Likelihood from the Gibbs Output. *Journal of the American Statistical Association* 1995; **90**:1313–1321.
25. Neal RM. Slice Sampling. *The Annals of Statistics* 2003; **31**:705–741.
26. Pons JMV, Granados A, Espinas JA, Borras JM, Martin I, Moreno V. Assessing Open Heart Surgery Mortality in Catalonia (Spain) through A Predictive Risk Model. *European Journal of Cardio-thoracic Surgery* 1997; **11**:415–423.
27. West M. Bayesian Model Monitoring. *Journal of Royal Statistical Society, Series B* 1986; **48**:70–78.
28. Bedrick EJ, Christensen R, Johnson W. A New Perspective on Priors for Generalized Linear Models. *Journal of the American Statistical Association* 1996; **91**:1450–1460.
29. Chaloner K, Duncan GT. Assessment of A Beta Prior Distribution: PM Elicitation. *The Statistician* 1983; **27**:174–180.
30. Gavasakar U. A Comparison of Two Elicitation Methods for A Prior Distribution for A Binomial Parameter. *Management Science* 1988; **34**:784–790.