ISQua®

*International Journal for Quality in Health Care*, 2015, 27(1), 31–36
doi: 10.1093/intqhc/mzu092
Advance Access Publication Date: 8 December 2014
Article

OXFORD

Article

# The impact of varying patient populations on the in-control performance of the risk-adjusted CUSUM chart

**WENMENG TIAN[1], HONGYUE SUN[1], XIANG ZHANG[2], and WILLIAM H. WOODALL[2]**

[1]Grado Department of Industrial and Systems Engineering, Virginia Tech, Blacksburg, VA 24061, USA, and [2]Department of Statistics, Virginia Tech, Blacksburg, VA 24061, USA

Address reprint requests to: William H. Woodall, Department of Statistics, Virginia Tech, Blacksburg, VA 24061, USA. Tel:+1-540-231-7792; Fax: +1-540-231-3863; E-mail: bwoodall@vt.edu

Accepted 13 October 2014

## Abstract

**Objective:** This research is designed to examine the impact of varying patient population distributions on the in-control performance of the risk-adjusted Bernoulli CUSUM chart.
**Design:** The in-control performance of the chart is compared based on sampling the Parsonnet scores with replacement from five realistic subsets of a given distribution.
**Settings:** Five patient mixes with different Parsonnet score distributions are created from a real patient population.
**Main Outcome Measures:** The outcome measures for this research are the in-control average run lengths (ARLs) given varying patient populations.
**Results:** Our simulation results show that the in-control ARLs of the risk-adjusted Bernoulli CUSUM chart with fixed control limits and a given risk-adjustment equation vary significantly for different patient population distributions, and the in-control ARLs decrease as the mean of the Parsonnet scores increases.
**Conclusions:** The simulation results imply that the control limits should vary based on the particular patient population of interest in order to control the in-control performance of the risk-adjusted Bernoulli CUSUM method.

**Key words:** average run length (ARL), heterogeneous population distributions, in-control performance, Parsonnet score, risk-adjusted CUSUM, statistical process control

## Introduction

The risk-adjusted Bernoulli CUSUM chart was developed by Steiner *et al.* [1], driven by the need for appropriately monitoring surgical performance in the presence of risk differences from patient to patient. With this chart, one adjusts for each patient's preoperative risk of surgical failure using a logistic regression model and then applies the likelihood ratio based scoring method to obtain the monitoring statistics. It has been shown that the risk-adjusted CUSUM chart is suitable for detecting improvement or deterioration in surgical performance when there is a mix of patients with varying preoperative risks. A general review of risk-adjusted charting was provided by Cook *et al.* [2]. If patient risks do not vary significantly, one can use the Bernoulli CUSUM chart employed by Lim *et al.* [3].

A number of practitioners and researchers have applied the risk-adjusted Bernoulli CUSUM method for monitoring clinical outcomes. Sherlaw-Johnson advised using the signal rule of the risk-adjusted CUSUM chart in the background with the perhaps more commonly used variable life-adjusted display (VLAD) [4]. This is the approach advocated by the Clinical Practice Improvement

Centre [5]. Sherlaw-Johnson *et al.* combined the risk-adjusted CUSUM chart and the rocket tail chart based on the VLAD for monitoring by developing a scheme in which CUSUM signals were superimposed onto the rocket tail plots [6]. In another application, Harris *et al.* retrospectively applied the risk-adjusted CUSUM method to the analysis of their medical center's experience with ruptured abdominal aortic aneurysms (RAAAs) while adjusting for the variability in patients' comorbidities and hemodynamic instability [7]. As another example, Novick *et al.* compared risk-adjusted and non-risk-adjusted CUSUM analyses of coronary artery bypass surgery outcomes and found out that the risk-adjusted CUSUM method was advantageous over non-risk-adjusted methods by not incorrectly signaling a deterioration in performance when preoperative patient risk was high [8]. Moore *et al.* used the risk-adjusted CUSUM method to assess shifts in the performance of an RAAA program over time [9]. As a final example, Coory *et al.* recommended use of the risk-adjusted CUSUM chart for monitoring administrative hospital data [10].

Setting appropriate control limits to get a desired in-control average run length (usually denoted as in-control ARL) is the prerequisite for evaluating the control chart performance. The run length is defined in our case as the number of Bernoulli trials (i.e. patients) observed until a signal is given by the control chart. The signal can be either a false alarm or an indication of the occurrence of a real change in the parameter of interest. Getting the same or very close in-control ARL values is the prerequisite for comparing the out-of-control performance of competing control charts. The selection of the control limits for various types of CUSUM schemes has been discussed by a number of authors, including Woodall [11] and Gan [12]. In the cardiac surgery example used by Steiner *et al.* [1], the control limits for the proposed CUSUM charts were set at a specified level to give a relatively large in-control ARL value given the patient population and the fitted logistic regression model used for risk adjustment.

The impact of varying patient population distributions on the performance of the risk-adjusted CUSUM chart has been recognized. Steiner *et al.* [13] discussed the change in ARL performance when changes occurred in the risk distribution by plotting the ARL versus odds ratio curves based on two extremely different patient distributions. One population consisted only of the lowest risk patients and the other consisted of only the highest risk patients. For these populations, the in-control ARLs varied from 100 to 1000. Rogers *et al.* [14] simply stated that the in-control ARL performance would change if the risk distribution changes over time. Loke and Gan [15] investigated the sensitivity of the risk-adjusted CUSUM charts to changes in the 'predisposed risk distribution' by using simulated beta distributions with different parameters to compare the in-control ARLs. For a set of somewhat arbitrarily selected beta distributions, they showed that the in-control ARL could be up to 13% lower than specified or up to 31% higher. They proposed a joint scheme to monitor the clinical failures and the risk distribution at the same time. In our study, we investigated the impact of varying risk factor distributions on the in-control performance of the risk-adjusted CUSUM chart based on more realistic populations from a case study application.

In our study, we examined the effect of varying patient population distributions on the in-control performance of the risk-adjusted CUSUM chart. The distribution of risk factors can vary significantly depending on the surgeon, the patient mix corresponding to various facilities, and even the same surgeon and facility over time. Given the risk model fitted during the preliminary Phase I and fixed control limits determined for a given population of patients, we would hope to have similar in-control performance in the on-going monitoring of Phase II for different patient mix distributions.

The remainder of our short paper is organized as follows. The risk-adjusted Bernoulli CUSUM procedure is introduced, and the various patient population risk factor distributions we consider are illustrated. Then we compare the in-control performance of the risk-adjusted CUSUM charts for different patient population distributions. Finally, our conclusions and a future research topic are discussed.

## Methods

### Risk-adjusted Bernoulli CUSUM chart

Due to the fact that the preoperative risk of mortality varies considerably from patient to patient in most surgical settings, Steiner *et al.* developed a monitoring approach in which one can adjust for each patient's preoperative risk of surgical failure [1]. It is thus referred to as the risk-adjusted Bernoulli CUSUM chart, where the CUSUM statistics for the $t$th patient can be written as follows:

$$C_t^+ = \max(0, C_{t-1}^+ + W_t^+), \quad t = 1, 2, 3, \ldots, \tag{1}$$

$$C_t^- = \min(0, C_{t-1}^- - W_t^-), \quad t = 1, 2, 3, \ldots, \tag{2}$$

where $C_0^+ = C_0^- = 0$, and $W_t^+$ and $W_t^-$ are the weights assigned to the $t$th patient which are determined based on a log-likelihood ratio depending on the observed surgical outcome. The surgical failure probability for each patient is estimated by evaluating the preoperative risk of each patient using a method such as a logistic regression model based on Parsonnet scores. (See Parsonnet *et al.* [16].)

Let $R$ denote the odds ratio corresponding to surgical failure and $P_t$ denote the probability of patient death within a 30-day period after the surgery determined from the risk-adjustment model. Then for patient $t$, the odds of failure are $RP_t$: $(1 - P_t)$ and the probability of failure is correspondingly$RP_t/(1 - P_t + RP_t)$. The risk-adjusted CUSUM chart is designed to monitor for a change from $R_0$ to $R_a$, where $R_0$ is usually set to 1 to reflect current surgical performance, $R_a$ to $R_a > R_0$ for detecting performance deterioration, and to $R_a < R_0$ to detect process improvement. Thus, the risk-adjusted CUSUM chart weights can be written as follows,

$$W_t = \begin{cases} \log\left(\dfrac{1 - P_t + R_0 P_t}{1 - P_t + R_a P_t}\right) & \text{if } y_t = 0, \\ \log\left[\dfrac{(1 - P_t + R_0 P_t) R_a}{(1 - P_t + R_a P_t) R_0}\right] & \text{if } y_t = 1. \end{cases} \tag{3}$$

where $y_t = 1$ if the patient $t$ dies within the 30-day period following the surgery, and $y_t = 0$ otherwise. When $R_a > 1$, Equation (3) yields values of $W_t^+$ in Equation (1). When $R_a < 0$, Equation (3) yields values of $W_t^-$ in Equation (2). The chart signals when $C_t^+ \geq h^+$ or $C_t^- \leq h^-$ which indicates that there has been either deterioration or improvement in the surgical performance, respectively. The control limits $h^+$ and $h^-$ are set to yield suitably large ARLs when there are no changes in the odds ratio of failure $R_0$.

### Varying patient populations

It is expected that the patient risk will vary due to the differences in personal characteristics such as age and the presence or absence of health risk factors such as hypertension or diabetes. As Steiner *et al.* [1] and many others have pointed out, risk adjustment is often necessary in health-care applications. In our work, we want to study how much the in-control performance of the risk-adjusted Bernoulli CUSUM chart depends on realistic patient populations. Before discussing the results of the simulation, we first introduce how the patient

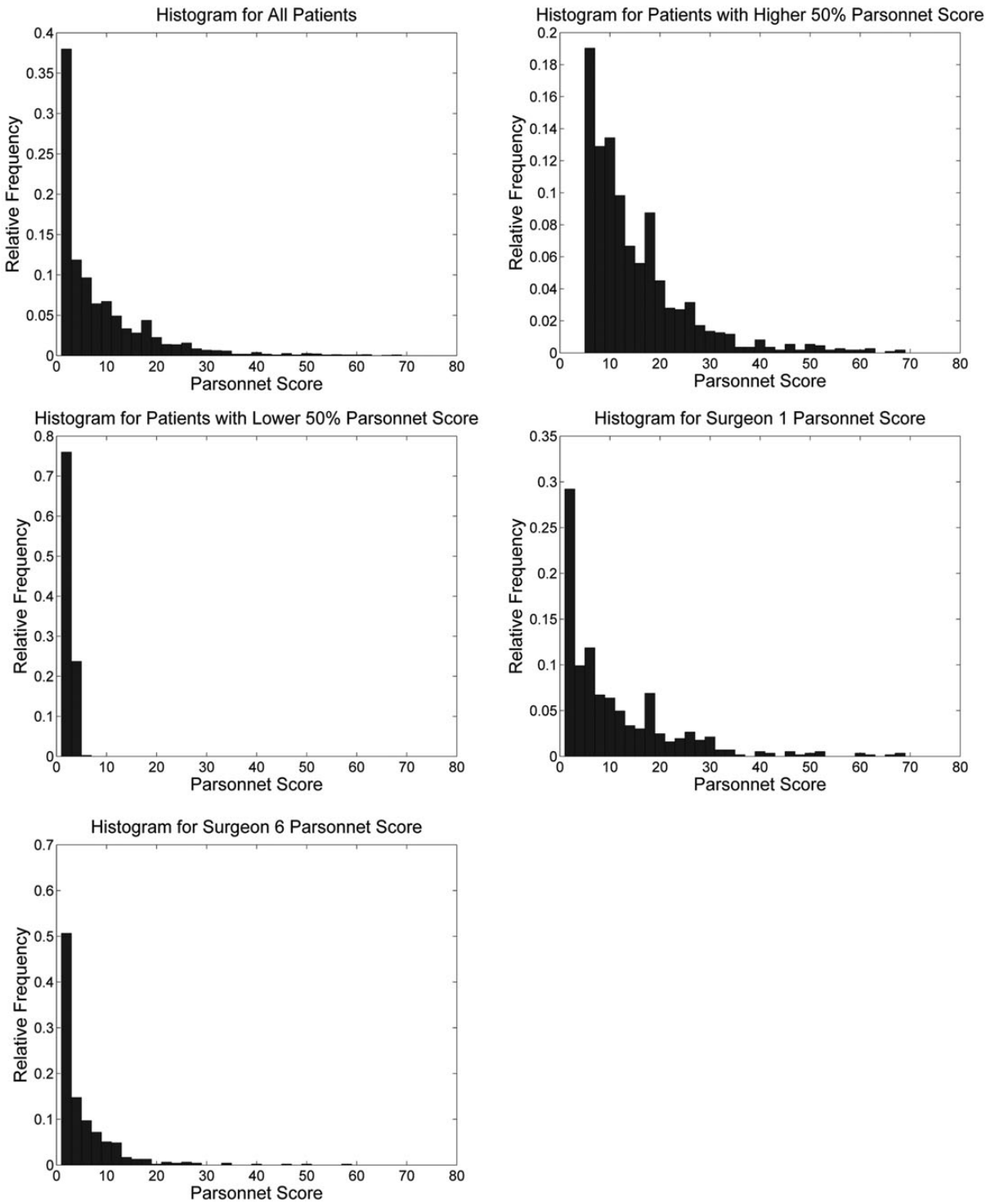


**Figure 1** Parsonnet score histograms for five populations.

distribution was varied using a dataset from a UK center for cardiac surgery (used in Steiner *et al.* [1]).

The dataset contains information on 6994 patients collected over a 7-year period from 1992 to 1998. For each patient, some surgeon information, the binary surgical outcome and a Parsonnet score are recorded. The first 2 years of data were used to fit the logistic regression model based on the Parsonnet score, which was determined based on the personal and health characteristics of each patient [1].

The histograms corresponding to the Parsonnet scores of all patients in the first 2 years, the lower 50% of these scores, the higher 50% of these scores, and the scores of the patients corresponding to the first and sixth surgeon are provided in Fig. 1. These distributions of risk factors vary considerably, which makes the distributions of the death rates of the patients ($P_t$'s) vary significantly as well, as illustrated in Fig. 2. We used these five distributions as our populations in our simulation study. We note that Jones and Steiner [17] sampled from

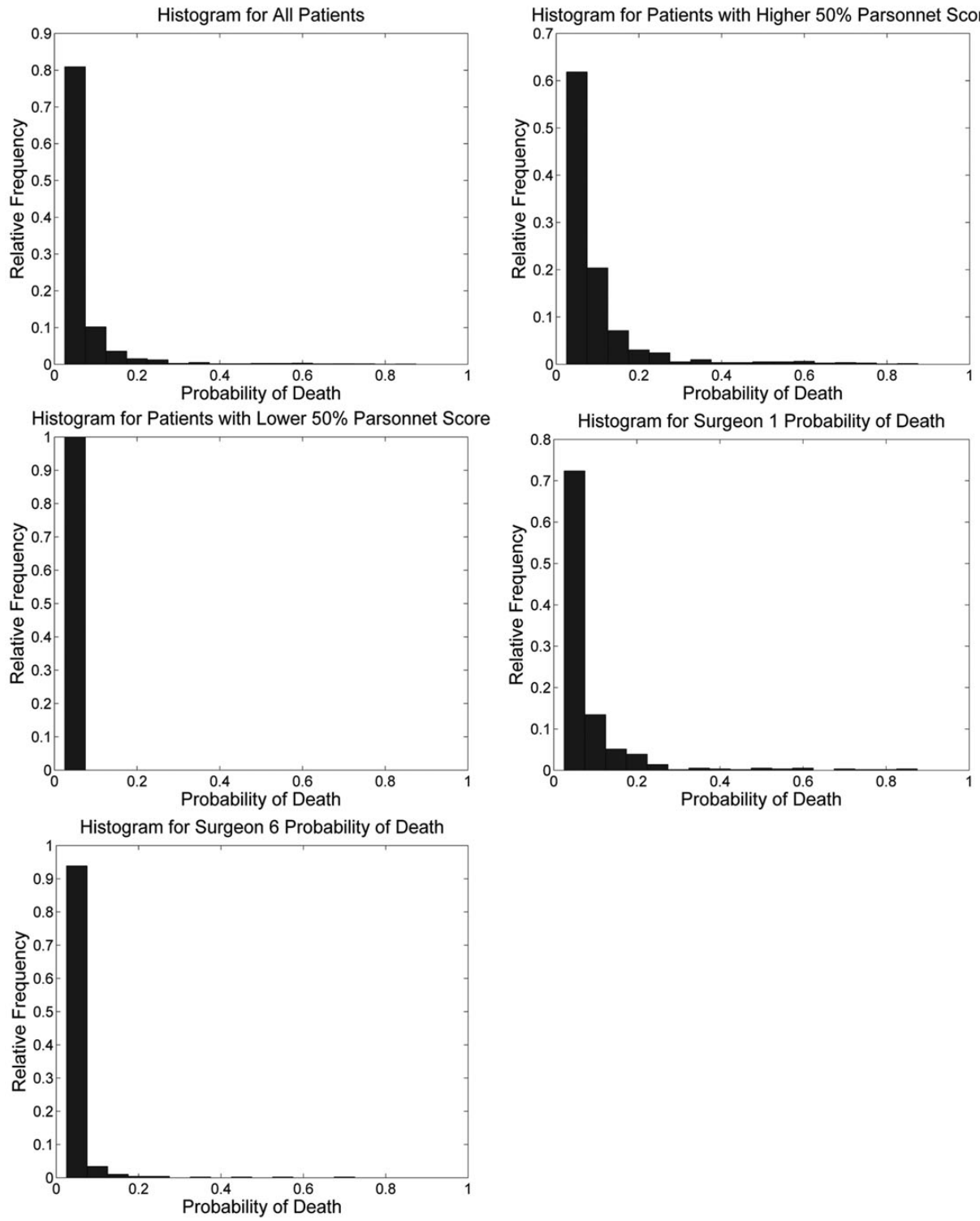


**Figure 2** Histograms of probability of death within 30 days following the surgery for five populations.

subsets of this dataset in a similar manner to study the effect of estimation error of the parameters of the risk-adjustment model on chart performance.

The mean values for the Parsonnet scores in each population are given in Table 1. One can see that the average Parsonnet score varies considerably. Based on the five patient populations we have described, we performed in-control performance comparisons based on the model estimated from all the surgeries in years 1992 and 1993, which Steiner *et al.* [1] found to be

$$\text{logit}(P_t) = -3.68 + 0.077\, X_t, \tag{4}$$

where $X_t$ represents the Parsonnet score of patient $t$, and $P_t$ is the probability of death within 30 days following surgery for this patient.

**Table 1** Population means and in-control ARL comparison

| Population | Mean Parsonnet score | Upper CUSUM ARL | Lower CUSUM ARL |
|---|---|---|---|
| All Phase I scores | 8.9026 | 7400.1 (73.6) | 6069.3 (59.6) |
| Lower 50% of scores | 2.0541 | 12 324.0 (120.1) | 11 014.0 (105.5) |
| Higher 50% of scores | 16.4813 | 4988.6 (50.9) | 3983.7 (39.8) |
| Surgeon 1 scores | 11.2513 | 6474.1 (63.9) | 5148.4 (50.2) |
| Surgeon 6 scores | 5.5591 | 9657.8 (94.2) | 8207.2 (78.7) |

The values in the parentheses are the standard errors of the estimated ARLs.

(Note that due to a typographical error, Steiner *et al.* [1] reported the coefficient as 0.77 instead of 0.077.)

To check the effect of the patient risk distribution on the performance of the risk-adjusted CUSUM chart proposed in Steiner *et al.* [1], scores were randomly sampled with replacement from the five distributions. The probabilities of death after the surgery were determined using Equation (4) based on the sampled scores. Then, the CUSUM statistics for both upper and lower CUSUM charts were obtained using Equations (1) and (2) based on Equation (3) with $R_a = 2$ for the upper CUSUM and $R_a = 0.5$ for the lower CUSUM after generating Bernoulli random variables with the probabilities from Equation (4).

## Results

If the varying Parsonnet score distributions have no significant impact on the chart performance, comparable in-control ARLs should be obtained for the varying distributions for the same control limits, $h^+ = 4.5$ and $h^- = -4.0$, which were determined by Steiner *et al.* [1] to yield in-control ARLs of roughly 9600 for each of the upper and lower charts. In our study, 10 000 run lengths were obtained by repeatedly sampling from each of the different populations with replacement, and the in-control ARLs were estimated. The results are summarized in Table 1. The values in the parentheses are the standard errors of the estimated ARLs.

First, we noted that the in-control ARL values based on all of the scores did not match the results of Steiner *et al.* [1]. According to Steiner (2013, personal communication), this was due to the limited number of states used in their Markov chain calculations. Increasing the number of states to 1000 led him to values of 7322.9 and 6242.7 for the upper and lower CUSUM charts, respectively, which are comparable with the corresponding values in Table 1. It can be seen that the in-control performance varies significantly with different distributions of the patient risk factor. In addition, there is an obvious decreasing trend in the in-control ARLs as the mean of the Parsonnet scores increases, as illustrated in Fig. 3. Furthermore, there is a somewhat constant difference between the ARLs of the upper and lower risk-adjusted CUSUM charts.

The decreasing trend in the in-control ARL as the average Parsonnet score increases indicates an increase in false alarm rates when the mean of the Parsonnet scores gets higher. Since we assumed in our simulations that the model held exactly, the trend cannot be due to any lack-of-fit in the risk-adjustment model, but due to the change of the risk factor distribution. The control chart limits could be changed to remove the relatively constant gap between the ARLs of the upper and lower CUSUM charts, if this is desired.

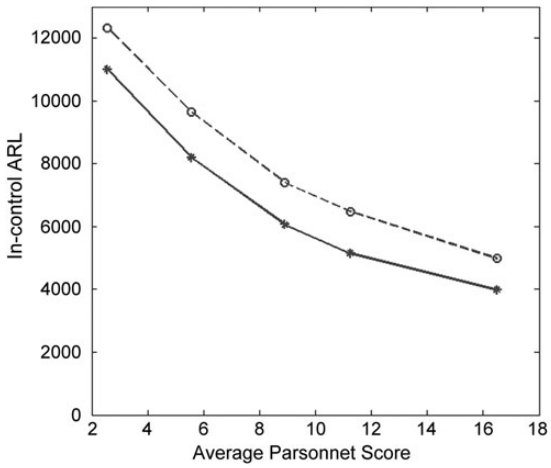


**Figure 3** The in-control ARLs of upper (circled) and lower (starred) risk-adjusted CUSUM charts given varying risk distributions.

## Discussion

In our paper, we examined the effect of varying patient risk factor distributions on the performance of the risk-adjusted CUSUM chart proposed by Steiner *et al.* [1]. With sampling only from subsets of the Phase I dataset with replacement, significantly different in-control ARLs were obtained. Also, the in-control ARLs decrease as the mean of the Parsonnet scores increases. The implication of our results is that even if the risk-adjustment method is accurate, one cannot control the in-control performance of the risk-adjusted Bernoulli CUSUM method without determining the control limits based on the particular assumed patient population of interest. We showed that the in-control ARL can vary by more than a factor of two under realistic scenarios.

One possible solution to this problem, which we plan to investigate, is to adjust the control limits over time, conditioning on the particular sequence of risk factors observed for the patients. Shen *et al.* applied this approach in a computationally intensive way to successfully monitor Poisson data with varying in-control mean values [18]. The primary idea is to maintain the conditional probability of a false alarm as a constant value given there is no false alarm for previous observations. One uses simulated observations on-line to set a dynamic probability control limit (DPCL) for each patient. By applying the DPCLs to the risk-adjusted Bernoulli CUSUM chart, one can obtain the desired in-control run length distribution regardless of the distribution of patient risks.

## References

1. Steiner SH, Cook RJ, Farewell VT *et al.* Monitoring surgical performance using risk-adjusted cumulative sum charts. *Biostatistics* 2000;**1**:441–52.
2. Cook DA, Duke G, Hart GK *et al.* Review of the application of risk-adjusted charts to analyze mortality outcomes in critical care. *Crit Care Resusc* 2008;**10**:239–51.
3. Lim TO, Soraya A, Ding LM *et al.* Assessing doctors' competence: application of CUSUM technique in monitoring doctors' performance. *Int J Qual Health Care* 2002;**14**:251–8.

4. Sherlaw-Johnson C. A method for detecting runs of good and bad clinical outcomes on variable life-adjusted display (VLAD) charts. *Health Care Manag Sci* 2005;**8**:61–5.
5. Clinical Practice Improvement Centre. *VLADS for Dummies*. Milton, Queensland: Wiley Publishing Australia Pty Ltd, 2008.
6. Sherlaw-Johnson C, Morton A, Robison MB *et al*. Real time monitoring of coronary care mortality: a comparison and combination of two monitoring tools. *Int J Cardiol* 2005;**100**:301–7.
7. Harris JR, Forbes TL, Steiner SH *et al*. Risk-adjusted analysis of early mortality after ruptured abdominal aortic aneurysm repair. *J Vasc Surg* 2005;**42**:387–91.
8. Novick RJ, Fox SA, Stitt LW *et al*. Direct comparison of risk-adjusted and non-risk-adjusted CUSUM analyses of coronary artery bypass surgery outcomes. *J Thorac Cardiovasc Surg* 2006;**132**:386–91.
9. Moore R, Nutley M, Cina CS *et al*. Improved survival after introduction of an emergency endovascular therapy protocol for ruptured abdominal aortic aneurysms. *J Vasc Surg* 2007;**45**:443–50.
10. Coory M, Duckett S, Sketcher-Baker K. Using control charts to monitor quality of hospital care with administrative data. *Int J Qual Health Care* 2008;**20**:31–9.
11. Woodall WH. The design of CUSUM quality control charts. *J Qual Technol* 1986;**18**:99–102.
12. Gan FF. An optimal design of CUSUM quality control charts. *J Qual Technol* 1991;**23**:279–86.
13. Steiner SH, Cook RJ, Farewell VT. Risk-adjusted monitoring of binary surgical outcomes. *Med Decis Making* 2001;**21**:163–9.
14. Rogers CA, Reeves BC, Caputo M *et al*. Control chart methods for monitoring cardiac surgical performance and their interpretation. *J Thorac Cardiov Surg* 2004;**128**:811–9.
15. Loke CK, Gan FF. Joint monitoring scheme for clinical failures and predisposed risks. *Qual Technol Quant Manag* 2012;**9**:3–21.
16. Parsonnet V, Dean D, Berstein AD. A method of uniform stratification of risk for evaluating the results of surgery in acquired adult heart disease. *Circulation* 1989;**79**:I3–12.
17. Jones MA, Steiner SH. Assessing the effect of estimation error on risk-adjusted CUSUM chart performance. *Int J Qual Health Care* 2012;**24**:176–81.
18. Shen X, Tsung F, Zou C *et al*. Monitoring Poisson count data with probability control limits when sample sizes are time-varying. *Nav Res Log* 2013;**60**:625–36.