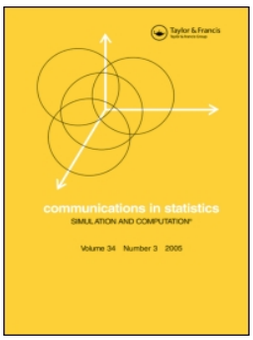

Taylor & Francis
Taylor & Francis Group

Communications in Statistics - Simulation and Computation

ISSN: 0361-0918 (Print) 1532-4141 (Online) Journal homepage: https://www.tandfonline.com/loi/lssp20

Performance of risk-adjusted cumulative sum charts when some assumptions are not met

A. Hussein, A. Kasem, S. Nkurunziza & S. Campostrini

To cite this article: A. Hussein, A. Kasem, S. Nkurunziza & S. Campostrini (2017) Performance of risk-adjusted cumulative sum charts when some assumptions are not met, Communications in Statistics - Simulation and Computation, 46:2, 823-830, DOI: 10.1080/03610918.2014.964805

To link to this article: https://doi.org/10.1080/03610918.2014.964805

Published online: 04 Nov 2016.

Submit your article to this journal

Article views: 105

View related articles

View Crossmark data

Citing articles: 3 View citing articles

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=lssp20

COMMUNICATIONS IN STATISTICS—SIMULATION AND COMPUTATION®
2017, VOL. 46, NO. 2, 823–830
http://dx.doi.org/10.1080/03610918.2014.964805

Taylor & Francis
Taylor & Francis Group

Performance of risk-adjusted cumulative sum charts when some assumptions are not met

A. Hussein[a], A. Kasem[a], S. Nkurunziza[a], and S. Campostrini[b]

[a]Department of Mathematics and Statistics, University of Windsor, Windsor, ON, Canada; [b]Dipartimento di Statistica, Universitá Cá Foscari di Venezia, San Giobbe, Venezia, Italy

ABSTRACT
Monitoring health care performance outcomes such as post-operative mortality rates has recently become more common, spurring new statistical methodologies designed for this purpose. One such methodology is the Risk-adjusted Cumulative Sum chart (RA-CUSUM) for monitoring binary outcomes such as mortality after cardiac surgery. When building RA-CUSUMs, independence and model correctness are assumed. We carry out a simulation study to examine the effect of violating these two assumptions on the chart's performance.

ARTICLE HISTORY
Received 15 July 2013
Accepted 9 September 2014

KEYWORDS
Correlated binary series; CUSUM charts; Logistic regression; Monitoring adverse events; Risk-adjustment

MATHEMATICS SUBJECT CLASSIFICATION
62P10

1. Introduction

Monitoring health care outcomes has become common in recent years with several new statistical methodologies being developed and introduced for this purpose.

The following four examples, which were also the first to trigger the current interest in the field, are often cited. The first is the report of Bristol Royal Infirmary Inquiry about the annual mortality rates for open-heart surgery on children under 1 year of age (Inquiry, 2001). The second is the case of the general practitioner Dr. Harold Shipman, who was convicted for murdering more than 200 of his patients (Inquiry, 2002). The third is monitoring risk of Down's Syndrome among newborn babies (Lie et al., 1993) and finally, the fourth is the monitoring of the incidence of congenital malformations after the Thalidomide Tragedy of 1960s. For a recent review of control charts for monitoring health care outcomes, we refer the reader to Woodall (2006).

When monitoring health care outcomes, it is often necessary to account for variations among the patient conditions that lead to differential prior risks at the time of treatment. That is, while monitoring the adverse outcomes of medical procedures, the heterogeneity of the baseline risk must be taken into account. This led to the notion of risk-adjusted charts in monitoring health care performance outcomes.

Various statistical process control (SPC) methodologies have been adapted to this new situation of health care performance monitoring. However, among the adopted methods, the Risk-adjusted Cumulative Sum Charts (RA-CUSUM) seem to be the most popular. Here, we will concentrate on the RA-CUSUMs, but the reader is referred to the review by Woodall

CONTACT A. Hussein ahussein@uwindsor.ca Department of Mathematics and Statistics, University of Windsor, 401 Sunset Avenue, Windsor, ON, N9B 3P4, Canada.
© 2017 Taylor & Francis Group, LLC

(2006) for other available methods. An early review on RA-CUSUM charts can be found in Grigg and Farewell (2004).

Originally, the RA-CUSUM charts were proposed in Lie et al. (1993) and later on formalized in the context of monitoring surgical outcomes by Steiner et al. (2000). The original format of the RA-CUSUM charts had several methodological issues, some of which have been resolved in recent literature. The most important among such issues were: ignoring the effect that autocorrelations may have on the chart performance; ignoring the effect that estimation of the baseline parameters from historical data could have on the chart performance; ignoring the effects of model misspecification; and high levels of false alarms (generally inherent in CUSUM charts). The effect of ignoring estimation of baseline parameters from historical datasets has been well-examined in recent literature. For instance, Jones and Steiner (2012) pointed out that, unless such estimation is properly accounted for, the Average Run Lengths (ARL) of the RA-CUSUM will deviate from their design values. A review of this issue and literature that attempted to rectify it can be found in Psarakis et al. (2014). Recently, Gombay et al. (2011) proposed a completely different approach to monitoring binary health care outcomes. This approach is based on truncated and risk-adjusted sequential testing procedures and is designed to reduce the probabilities of false alarms associated with the CUSUM charts. The same approach was later extended to accommodate autocorrelations and retrospective surveillance of binary outcomes (see Fokianos et al., 2013).

In this article, we concentrate only on examining the effect of autocorrelations and model misspecification on the performance of the RA-CUSUM. In Section 2, we give a brief introduction to the RA-CUSUM method. In Section 3, we use Monte Carlo simulations to assess the effect of autocorrelations and model misspecifications on the Average Run Lengths (ARL) of the RA-CUSUM. In Section 4, we give some conclusions.

2. Risk-adjusted CUSUM charts and some of its assumptions

The risk-adjusted CUSUM (RA-CUSUM) was first proposed in Lie et al. (1993) who applied it to monitoring Down's syndrome. The method was later on formalized and applied in the context of monitoring surgical outcomes by Steiner et al. (2000).

Let us define

$$\pi_t = g(\eta, \mathbf{X_t})$$

as the probability of adverse event for the tth patient, where $\boldsymbol{X_t}$ is a $p \times 1$ vector, $\boldsymbol{X_t} = (X_{t1}, X_{t2}, \ldots, X_{tp})^T$, representing the risk factors for patient t and η is a vector of parameters, which may represent the effect of the various risk factors. The function $g(.)$ may be determined apriori using a rating method, for example, Parsonnet risk score (Parsonnet et al., 1989) or may be based on a logistic regression model fitted to sample data. Often, the function is just the inverse of the $\log(1/(1-x))$ function connecting the patient covariates to the probability of adverse event. These models are often called baseline or reference risk models.

Let the null and alternative, H_0 and H_A, respectively, represent the in-control and out-of-control processes. These hypotheses can be based on odds ratio because each patient has a different baseline risk level. For a given estimated risk of adverse event equal to π_t, the odds of adverse event is $\frac{\pi_t}{1-\pi_t}$. The risk-adjusted CUSUM allows for varying π_t of the patient population by sequentially testing the hypotheses,

$$H_0 : \frac{\pi_t^0/(1-\pi_t^0)}{\pi_t/(1-\pi_t)} = R_0 \tag{1}$$

versus

$$H_A : \frac{\pi_t^1/(1-\pi_t^1)}{\pi_t/(1-\pi_t)} = R_A, \tag{2}$$

where π_t^0 is the probability of an adverse outcome for an in-control process and R_0 is the ratio of the odds of an adverse outcome for an in-control process to the odds of an adverse outcome after risk-adjustment for patient t. Similarly, π_t^1 is the probability of an adverse outcome for an out-of-control process with R_A being the ratio of the odds of an adverse outcome for an out-of-control process to the odds of an adverse outcome after risk-adjustment for patient t. Usually, $R_A > R_0$ indicates increase in the failure rate of the treatment. If the estimated risk π_t is based on the current conditions, as described in Steiner et al. (2000), then we may set $R_0 = 1$. Such hypotheses can also be formulated in terms of π_t^0 and π_t^1. In fact, we can see that

$$\pi_t^0 = \frac{R_0\pi_t}{1-\pi_t+R_0\pi_t}$$

and

$$\pi_t^1 = \frac{R_A\pi_t}{1-\pi_t+R_A\pi_t}.$$

Thus, the likelihood ratio for the RA-CUSUM scheme becomes

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

where $y_t = 1$ if adverse event took place for the tth patient and zero otherwise.

The RA-CUSUM procedure is to monitor the quantity

$$Z_t = Max\left[0, Z_{t-1} + W_t\right], \tag{4}$$

where $Z_0 = 0$ and W_t is the sample likelihood ratio or score assigned to the tth patient, as defined above. Such a procedure is designed to detect poor performance of surgeons, as it is restricted to non-negative values. If the quantity Z_t crosses a positive threshold $h > 0$, that is, $Z_t > h$, then the process is signaled to be out-of-control, whereas if $Z_t = 0$, then the monitoring procedure is re-started. For detecting improved performances, Steiner et al. (2000) suggested using

$$Z_t = \min\left[0, Z_{t-1} - W_t\right]. \tag{5}$$

CUSUM charts are designed to monitor the responses sequentially until there is a deterioration in the process. In general, the CUSUM control chart will eventually give an out-of-control signal that might be a false alarm. The average number of observations required before an alarm (whether false or true) is known as Average Run Length (ARL) of the CUSUM. An appropriate value of threshold (control limit) h depends on the desired levels of ARL under H_0 and H_A. It is desired that the ARL should be long enough under H_0 since in this situation the process is in-control and any alarm would be a false alarm. Likewise, the ARL should be as short as possible under H_A as in that case the alarm would be signaling a legitimate out-of-control situation. The ARL of the RA-CUSUM under H_0 and H_A are considered analogous to the Type I and Type II error rates of the traditional statistical test, respectively. However, unlike these familiar types of errors, there is no standard criteria to decide how large or small ARLs should be.

The probability π_t of adverse event for the tth patient is a function of the patient's covariates. In cardiac surgery, a logistic model based on Parsonnet score (see Parsonnet et al., 1989) is used so that

$$\log\left(\frac{\pi_t}{1-\pi_t}\right) = \alpha + \beta x_t,$$

where x_t is the Parsonnet score. The parameters α and β are estimated from historical datasets.

A concern with the current RA-CUSUM charts for monitoring health outcomes is the fact that the sequence of observations y_t is assumed to be uncorrelated when in reality this is unlikely. It is likely that if we monitor outcomes from a sequence of patients operated by the same surgeon, the observations thereof will be autocorrelated. Such correlations, if not appropriately accounted for, can result in misleading charts. The second major concern in the RA-CUSUM procedures is the fact that the baseline model, often based on logistic regression with Parsonnet score as the sole covariate, may be incorrect. Although Parsonnet scores summarize efficiently the patient's prior risks, there might be some other important covariates that are not accounted for in the model. Not accounting for important covariates would lead to a misspecified baseline model and hence misleading results when monitoring the data. As far as the authors are aware of, there are no studies assessing robustness of the RA-CUSUM charts against baseline model misspecification.

In the next section, we will run some Monte Carlo simulations to examine the extent to which various autocorrelations as well as baseline model misspecification would impact on the performance of the RA-CUSUM charts.

3. Simulation studies

Throughout these simulations, we assume that the patient model is built from a logistic regression with intercept and Parsonnet score as the only covariate, that is, with parameters (α, β). Following the cardiac surgery data example used by Steiner et al. (2000), we assume that $\alpha = -3.68$ and $\beta = 0.077$.

In order to examine the effect of autocorrelations on the chart's performance, we have to generate correlated binary data. In the cardiac model considered here, the probabilities of adverse events depend on the patient's Parsonnet score that leads to differential marginal probabilities and hence to non-stationarity. Generating a correlated non-stationary binary sequence is generally a difficult task since the marginal probabilities determine the correlation structure of the sequence. The most flexible way of generating non-stationary correlated Bernoulli is perhaps the one proposed by Oman and Zucker (2001). In this reference, algorithms for generating correlations close to the familiar intraclass, AR(1) and MA(1) correlations are provided. We point out here that intraclass correlation means that every pair of observations has the same fixed correlation coefficient. In the following, we re-phrase the Oman-Zucker algorithm for our current context.

- Generate Parsonnet scores, x_t, from Poisson distribution with $\lambda = 25$
- Compute

$$\pi_t = \frac{R\pi_t^0}{1-\pi_t^0+R\pi_t^0}, \tag{6}$$

where

$$\pi_t^0 = \frac{\exp(-3.68+0.077\times x_t)}{1+\exp(-3.68+0.077\times x_t)} \tag{7}$$

is the baseline rate for the tth patient and R is either the odds ratio under H_0, $R_0 = 1$ or the odds ratio under the alternative of doubling the odds ratio to $R_A = 2$.

- Generate γ_t as the standard normal quantile to the probability π_t, that is, $\Phi^{-1}(\pi_t) = \gamma_t$.
- Generate U_t, independent Bernoulli random variables with probability of success θ (a pre-fixed correlation parameter).
- Generate ε_t for $t = 0, 1, 2, \ldots,$ independent standard normal variates, independent of the U_t.
- For $t = 1, 2, 3, \ldots,$ set

$$Y_t = \begin{cases} 1 & \text{if } z_t \leq \gamma_t \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

In order to generate Y_t with intraclass, AR(1) or MA(1), we choose $z_t = U_t\varepsilon_0 + (1 - U_t)\varepsilon_t$, $z_t = U_t\varepsilon_{t-1} + (1 - U_t)\varepsilon_t$, or $z_1 \sim N(0, 1)$, $z_{t+1} = U_{t+1}z_t + (1 - U_{t+1})\varepsilon_{t+1}$, respectively. Here $z_1 \sim N(0, 1)$ is a random number generated from the standard normal distribution.

In the case of the intrclass correlation, Oman and Zucker (2001) showed that the pairwise correlations between any two of the binary sequence is proportional to θ^2 with a fixed proportionality coefficient. Thus, resulting in pairwise constant correlation coefficients among the Y_t. Similarly, in the AR(1) and MA(1) cases, the correlation between Y_t and Y_s is proportional to $I\{|t - s| = 1\}\theta(1 - \theta)$, $\theta^{|t-s|}$, respectively. Therefore, θ is obviously the parameter controlling the strength of the correlations.

The results for various values of θ and threshold values h are summarized in Tables 1–3. Each entry in these tables is obtained by 5,000 Monte Carlo runs except the results of Table 1 where, due to the prohibitively large ARLs, only 1,000 Monte Carlo runs were performed. Notice that the benchmark for comparison is the first row of each table that represents the uncorrelated case (i.e., $\theta = 0$).

The second set of Monte Carlo simulations were carried out to examine the impact of model misspecification on the ARLs of the RA-CUSUM. To this end, we generate the patient's adverse event rate from

$$\pi_t^0 = \frac{\exp(-3.68 + 0.077 \times x_t + \epsilon_t)}{1 + \exp(-3.68 + 0.077 \times x_t + \epsilon_t)}. \tag{9}$$

That is, we introduce a normally distributed random effect, $\epsilon_t \sim N(0, \sigma)$ with a pre-fixed standard deviation σ, in the patient's logistic regression baseline model. The null and the alternative hypotheses remain as above, $R_0 = 1$, $R_A = 2$. The purpose of including random effects

Table 1. Simulated average run lengths, ARL_0 under H_0 and ARL_1 under H_A for various RA-CUSUM charts (different h) and various intraclass-type correlation parameters, θ. Extremely large ARLs are replaced by ∞.

	$h = 2.5$		$h = 3$		$h = 3.5$		$h = 4$	
θ	ARL_0	ARL_1	ARL_0	ARL_1	ARL_0	ARL_1	ARL_0	ARL_1
0.00	331.42	44.99	606.34	57.09	1134.23	65.42	1924.68	84.56
0.05	415.37	46.06	721.57	56.89	1361.58	70.04	2353.28	87.18
0.10	512.50	54.19	951.35	64.84	2044.41	79.44	3605.89	107.01
0.15	641.92	56.83	1349.41	74.12	2740.08	99.65	6269.81	117.70
0.20	853.92	71.08	1902.34	96.60	4966.50	116.10	11197.34	126.96
0.25	1209.18	77.10	2921.55	120.53	7870.21	149.11	20930.14	191.39
0.30	1996.18	98.69	4918.72	160.10	13957.49	211.84	34908.76	286.58
0.35	2469.41	128.60	8359.35	226.94	27790.83	321.02	78422.36	462.50
0.40	4118.92	176.72	15179.22	331.34	48092.73	521.79	∞	985.17
0.45	6613.88	266.91	28034.10	537.03	110260.33	1016.52	∞	1719.45

Table 2. Simulated average run lengths, ARL_0 under H_0 and ARL_1 under H_A for various RA-CUSUM charts (different h) and various AR(1)-type correlation parameters, θ.

	$h = 2.5$		$h = 2.75$		$h = 3.0$		$h = 3.25$	
θ	ARL_0	ARL_1	ARL_0	ARL_1	ARL_0	ARL_1	ARL_0	ARL_1
0.00	347.24	45.32	473.22	51.86	618.14	57.78	828.10	62.59
0.05	291.41	45.44	379.11	51.31	513.21	55.79	662.48	62.72
0.10	247.59	44.56	330.03	49.68	427.82	54.97	544.70	62.11
0.15	219.99	43.84	280.91	48.67	348.40	54.29	450.14	59.90
0.20	190.47	42.96	240.96	48.02	311.96	54.65	381.05	59.62
0.25	165.81	41.88	210.01	48.00	261.10	53.70	323.36	59.95
0.30	146.90	41.95	187.62	47.27	226.43	51.94	281.49	56.41
0.35	133.01	40.75	165.14	47.23	204.16	50.31	238.76	55.93
0.40	120.93	39.98	148.50	44.98	177.02	49.96	209.66	54.79
0.45	110.75	39.11	129.60	43.36	161.97	48.70	190.41	53.61
	$h = 3.5$		$h = 3.75$		$h = 4.0$		$h = 4.25$	
0.00	1060.33	69.92	1431.15	74.63	1854.67	81.00	2509.37	87.23
0.05	869.58	68.32	1118.92	73.81	1418.27	80.67	1832.01	85.73
0.10	698.06	67.65	886.99	72.76	1103.73	79.40	1391.67	85.11
0.15	565.02	67.85	713.89	72.04	871.70	79.03	1089.99	84.31
0.20	473.59	65.43	583.10	71.06	722.57	77.26	875.15	82.67
0.25	394.73	64.59	480.07	70.75	581.05	75.78	702.47	82.63
0.30	335.62	63.07	413.38	68.77	491.06	73.27	584.94	80.75
0.35	289.14	63.23	353.81	66.68	391.55	74.08	477.63	79.20
0.40	252.69	60.84	292.79	65.58	344.33	72.13	403.98	76.35
0.45	215.49	60.67	255.80	64.53	292.99	70.66	341.07	73.69

Table 3. Simulated average run lengths, ARL_0 under H_0 and ARL_1 under H_A for various RA-CUSUM charts (different h) and various MA(1)-type correlation parameters, θ.

	$h = 2.5$		$h = 2.75$		$h = 3.0$		$h = 3.25$	
θ	ARL_0	ARL_1	ARL_0	ARL_1	ARL_0	ARL_1	ARL_0	ARL_1
0.00	348.34	46.22	462.55	51.53	619.82	56.93	828.12	63.15
0.05	292.91	44.95	400.18	50.82	525.21	56.98	678.41	61.50
0.10	260.28	44.36	349.94	50.06	454.57	56.02	592.83	60.38
0.15	240.94	43.59	312.73	50.49	397.48	56.05	525.41	60.86
0.20	221.76	42.78	291.66	49.40	375.54	55.24	469.32	60.89
0.25	217.12	43.21	276.23	48.54	348.77	54.08	440.66	60.86
0.30	203.79	42.75	260.98	48.31	335.93	54.63	418.45	61.18
0.35	195.29	42.32	256.80	47.63	322.47	53.38	402.43	58.93
0.40	190.83	43.15	246.22	48.28	306.77	54.62	389.38	59.22
0.45	187.29	42.43	245.70	48.50	303.00	53.33	389.83	59.85
	$h = 3.5$		$h = 3.75$		$h = 4.0$		$h = 4.25$	
0.00	1096.63	69.37	1403.00	74.67	1823.00	81.91	2403.77	87.33
0.05	877.01	69.56	1132.63	74.81	1485.28	82.01	1880.48	86.59
0.10	748.48	68.84	964.71	74.22	1218.75	80.45	1529.78	85.36
0.15	651.98	65.70	816.68	73.50	1033.22	80.82	1336.68	83.90
0.20	595.58	67.37	745.65	71.95	928.78	78.50	1165.22	84.36
0.25	545.06	66.14	696.34	71.97	848.80	77.86	1075.06	84.03
0.30	527.88	65.65	646.85	70.59	801.51	76.79	998.21	84.57
0.35	503.21	64.94	619.19	70.74	755.94	77.89	962.81	82.30
0.40	482.60	65.64	587.25	71.22	738.94	76.27	911.70	82.69
0.45	484.00	65.04	586.64	71.14	720.15	77.57	901.87	82.02

Table 4. Simulated average run lengths, ARL_0 under H_0 and ARL_1 under H_A for various RA-CUSUM charts (varying h). The binary series is generated from a misspecified baseline model with $\text{logit}(p) = \alpha + \beta x + \epsilon$, where ϵ is a mean-zero normal random effect with standard deviation σ.

	$h = 2.25$		$h = 2.5$		$h = 2.75$		$h = 3.0$	
$N(0, \sigma)$	ARL_0	ARL_1	ARL_0	ARL_1	ARL_0	ARL_1	ARL_0	ARL_1
$N(0, 0.2)$	251.24	39.87	341.91	45.51	460.17	51.89	622.08	57.23
$N(0, 0.4)$	253.10	40.60	346.52	46.15	463.28	52.05	632.34	57.50
$N(0, 0.5)$	256.59	40.64	352.98	46.90	475.35	52.68	622.22	58.72
$N(0, 0.6)$	255.04	41.40	342.94	46.83	479.03	52.46	616.02	58.46
$N(0, 0.8)$	262.65	41.56	347.09	47.97	480.52	53.40	640.98	60.42
$N(0, 1.0)$	264.89	42.46	360.61	48.59	494.24	55.35	654.55	60.93
$N(0, 1.25)$	274.67	44.68	362.11	49.35	497.09	57.10	661.36	63.19
$N(0, 1.5)$	287.20	46.06	388.89	53.71	527.72	59.52	692.73	66.42
$N(0, 1.75)$	297.06	49.06	402.29	55.23	540.88	63.35	727.28	70.35
$N(0, 2.0)$	310.98	51.88	422.06	59.25	561.08	67.46	755.07	73.76
$N(0, \sigma)$	$h = 3.25$		$h = 3.5$		$h = 3.75$		$h = 4.0$	
$N(0, 0.2)$	826.27	63.66	1092.46	68.59	1444.74	75.92	1869.32	82.36
$N(0, 0.4)$	820.52	63.65	1100.95	70.98	1397.00	76.52	1885.17	82.16
$N(0, 0.5)$	824.30	64.52	1082.89	70.86	1460.77	75.30	1866.44	82.76
$N(0, 0.6)$	809.01	65.16	1103.37	70.32	1425.69	78.04	1846.81	83.02
$N(0, 0.8)$	824.60	65.70	1097.23	72.70	1435.80	78.77	1880.61	83.18
$N(0, 1.0)$	853.05	66.10	1119.60	73.38	1505.92	80.39	1997.51	87.72
$N(0, 1.25)$	895.43	68.91	1171.98	75.97	1535.31	83.63	2011.42	89.65
$N(0, 1.5)$	935.01	73.27	1207.27	79.41	1595.88	86.84	2044.45	93.35
$N(0, 1.75)$	955.62	77.69	1282.29	84.24	1636.20	90.18	2170.91	98.78
$N(0, 2.0)$	1018.35	81.70	1352.22	81.71	1740.99	95.38	2352.00	105.09

is to examine what would happen if, besides the Parsonnet score, there were other important variables that have not been included in the model. The importance of the missed out covariates increase as the standard deviation of the random effects increases. The results of these simulations are reported in Table 4 for various random effect standard deviations σ and several threshold values h.

Table 1 shows that, in the case of intraclass correlations, both ARL_0 and ARL_1 increase as the strength of the correlations increase. That is, as the parameter θ increases. This can be seen by comparing the rows in Table 1 to the first row when there is no intraclass correlation (row with $\theta = 0$). This simply means that high intraclass correlations lead to delayed signals of poor performance. Also from Tables 2 and 3, we see that under correlations of types AR(1) and MA(1), the ARL_1 remains stable at design values (i.e., its values under $\theta = 0$), while ARL_0 drops substantially as correlations become stronger. This leads to too many false alarms in the surveillance process.

Regarding the problem of model misspecification, Table 4 indicates that when σ increases and hence the importance of the missed out covariates increase, both ARL_0 and ARL_1 increase slightly. This may again lead to less false alarms but a delay in giving genuine out-of-control alarms, as was the case of the intraclass correlations. However, it seems that in general the effect of model misspecification is not severe.

4. Conclusion

In this article, we considered the problem of monitoring adverse events in health care settings. Typical examples are the monitoring of mortality rates after cardiac surgery, monitoring of infectious diseases, monitoring Down's syndrome, influenza-like illnesses, and surveillance against bioterrorism. Recently, there has been a growing interest in developing

statistical methods that can satisfy the demands and the diversity of the problems present in health care outcomes surveillance. In this article, we focused on the so called Risk-adjusted CUSUM methodologies (RA-CUSUM), which are applied in particular to the monitoring of cardiac surgery outcomes. We gave some literature review on RA-CUSUM and pointed out that correlations and possible baseline model misspecifications are still some of the outstanding issues with the RA-CUSUM charts. By using Monte Carlo simulations, we examined the effect that ignoring such issues could have on the performance of the charts. We found out that if autocorrelations are present in the binary series being monitored and such autocorrelations are ignored, the average run lengths of the charts can deviate greatly from their design values. This is more severe in the case of intraclass correlations and less so in the AR(1) and MA(1) types of correlations. Such deviations lead to either too many false alarms or delayed genuine alarms. On the other hand, we found that the impact of model misspecification on the run lengths is not severe.

References

Fokianos, K., Gombay, E., Hussein, A. (2013). Retrospective change detection for binary time series. *Journal of Statistical Planning and Inferences* 145:102–112.

Gombay, E., Hussein, A., Steiner, S. (2011). Monitoring binary outcomes using risk-adjusted charts: A comparative study. *Statistics in Medicine* 30(23):2815–2826.

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

Inquiry. (2001). *BRI Inquiry Panel. Learning from Bristol: The Report of the Public Inquiry into Children's Heart Surgery at the Royal Infirmary 1984-1995*. London, UK: The Stationery Office. Available at: http://wwwbristol-inquiryorguk/final_report/.

Inquiry. (2002). *Shipman Inquiry. The First Report*. London, UK: The Stationery Office. Available at: http://wwwthe-shipman-inquiryorguk/reprt/asp.

Jones, M., Steiner, S. (2012). Assessing the effect of estimation error on risk-adjusted CUSUM chart performance. *International Journal of Quality in Health Care* 24(2):176–181.

Lie, R., Heuch, I., Irgens, L. (1993). A new sequential procedure for surveillance of downs-syndrome. *Statistics in Medicine* 12(1):13–25.

Oman, S. D., Zucker, D. M. (2001). Modelling and generating correlated binary variables. *Biometrika* 88(1):287–290.

Parsonnet, V., Dean, D., Bernstein, A. (1989). A method of uniform stratification of risk for evaluating the results of surgery in acquired adult heart-disease. *Circulation* 79(6, Suppl. S):3–12.

Psarakis, S., Vyniou, A. K., Castagliola, P. (2014). Some recent developments on the effects of parameter estimation on control charts. *Quality and Reliability Engineering International* 30(8):1113–1129.

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.

Woodall, W. (2006). The use of control charts in health-care and public-health surveillance. *Journal of Quality Technology* 38(2):89–104.