International Journal for Quality in Health Care 2012; **Volume 24, Number 2:** pp. 176–181
Advance Access Publication: 21 December 2011
10.1093/intqhc/mzr082

Assessing the effect of estimation error on risk-adjusted CUSUM chart performance

MARK A. JONES[1] AND STEFAN H. STEINER[2]

[1]Centre for Healthcare Related Infection Surveillance and Prevention (CHRISP), Queensland Health, Herston, QLD 4006, Australia, and [2]Department of Statistics and Actuarial Sciences, University of Waterloo, Waterloo, ON, Canada N2L 3G1

Address reprint requests to: Mark Jones, Centre for Healthcare Related Infection Surveillance and Prevention (CHRISP), Queensland Health, 15 Butterfield St, Herston, QLD 4006, Australia. Tel: +61-7-33289774; Fax: +61-7-33289769; E-mail: m.jones@sph.uq.edu.au

Accepted for publication 31 October 2011

Abstract

Background. Risk-adjusted control charts have become popular for monitoring processes that involve the management and treatment of patients in hospitals or other healthcare institutions. However, to date, the effect of estimation error on risk-adjusted control charts has not been studied.

Methods. We studied the effect of estimation error on risk-adjusted binary cumulative sum (CUSUM) performance using actual and simulated data on patients undergoing coronary artery bypass surgery and assessed for mortality up to 30 days post-surgery. The effect of estimation error was indicated by the variability of the 'true' average run lengths (ARLs) obtained using repeated sampling of the observed data under various realistic scenarios.

Results. Results showed that estimation error can have a substantial effect on risk-adjusted CUSUM chart performance in terms of variation of true ARLs. Moreover, the performance was highly dependent on the number of events used to derive the control chart parameters and the specified ARL for an in-control process (ARL_0). However, the results suggest that it is the uncertainty in the overall adverse event rate that is the main component of estimation error.

Conclusions. When designing a control chart, the effect of estimation error could be taken into account by generating a number of bootstrap samples of the available Phase I data and then determining the control limit needed to obtain an ARL_0 of a pre-specified level 95% of the time. If limited Phase I data are available, it may be advisable to continue to update model parameters even after prospective patient monitoring is implemented.

Keywords: CUSUM chart, estimation error, process monitoring, risk adjustment, statistical process control

Introduction

Control charts have been used for quality monitoring of industrial processes for many years. At the design stage, typically a stable (in-control) process is sampled over a period of time and the observed data are used to estimate parameters so that the distribution of the stable process can be determined. This stage of the monitoring process is called 'Phase I' data collection [1]. This estimated distribution is used to design a control chart to monitor the process to ensure it remains in control or conversely to provide a timely signal when it goes out of control. An obvious symptom of an out-of-control process is a change in the mean and/or variation. Clearly, the performance of the control chart will be influenced by the precision of the parameter estimates used to design the chart. In this (industrial) context, there have been a number of studies that have assessed the effect of parameter estimation on control chart performance. For a review of these studies, see ref. [2].

In recent years, control charts have become popular for monitoring processes that involve the management and treatment of patients in hospitals or other healthcare institutions. Unlike most industrial processes where each unit of observation is intended to be the same, patients are heterogeneous and this needs to be considered when designing control charts. This can be achieved by estimating and taking into account individual patient risks in the monitoring process. Control charts that implement this procedure are often referred to as risk-adjusted control charts. Steiner *et al.* [3] developed a risk-adjusted cumulative sum (CUSUM) control chart appropriate for binary (adverse) events, which are the type of events (along with counts and rates) most commonly encountered in the health context. The risk adjustment is achieved by estimating each patient's risk of an adverse event, e.g. using a logistic regression model and updating the CUSUM chart with a likelihood-based scoring method. These scores are used, in conjunction with actual patient outcomes, to produce weights for inclusion in a CUSUM control chart. As in the industrial context, the performance

of risk-adjusted control charts will be influenced by the precision of the estimated parameters used to derive the patient risk scores. Moreover, the added complexity of risk adjustment means that the previous studies of the effect of parameter estimation on control chart performance may not be relevant.

In this healthcare context, estimation error is the error associated with estimating the risk of an adverse event for each patient as well as estimating the proportion of patients in each risk group (the patient mix). In a simplified version of the real example, we use later in this study, there could be three risk groups in a population of around 7000 patients undergoing coronary artery bypass surgery: low, medium and high, with risk of 30-day mortality of 3, 9 and 21%, respectively. In addition, suppose the proportion of patients in each risk group is 62, 25 and 13%, respectively. Estimation error occurs because population parameters are estimated using a sample which is subject to sampling error due to random variation.

In this article, we assess the effect of estimation error on the performance of risk-adjusted CUSUM control charts. In addition, the effect of important variables such as adverse event rate, patient variability and desired average run length for an in-control process (ARL_0) on estimation error is investigated. In one scenario, we assume patients having a uniform adverse event risk, i.e. no risk adjustment is required. While Steiner *et al.* [3] reported on a small simulation study showing the sensitivity of their risk-adjusted CUSUM procedure to the initial estimate of patient mix and regression parameters, to our knowledge this is the first time the effect of estimation error has been systematically studied in the case of a simple binary CUSUM. We assume that the risk-adjustment model is correct, i.e. there is no confounding bias and important explanatory variables are not missing from the model. Assessing the performance of risk-adjusted control charts with modeling error is beyond the scope of this study.

Methods

To set up risk-adjusted control charts, we need estimates of:

(i) the patient mix, i.e. the (joint) distribution of the covariates and
(ii) the risk-adjustment model, e.g. a logistic regression model fit for a binary outcome.

We assume a single continuous covariate, e.g. Parsonnet score (predictive score for acquired adult heart surgery) [4], denoted by z. The patient mix (distribution of z) is denoted by $f(z)$ and the risk-adjustment model:

$$\text{Logit}\ (y) = \beta_0 + \beta_1 z. \qquad (1)$$

The risk-adjusted CUSUM procedure [3] involves sequentially monitoring:

$$X_t = \max\ (0, X_{t-1} + W_t), \quad t = 1, 2, 3, \ldots, \qquad (2)$$

where $X_0=0$ and W_t is the score for the tth patient, which is defined as

$$\log\left[\frac{(1 - p_t + R_0 p_t)R_A}{(1 - p_t + R_A p_t)R_0}\right] \quad \text{if } y_t = 1,$$

or

$$\log\left[\frac{(1 - p_t + R_0 p_t)}{(1 - p_t + R_A p_t)}\right] \quad \text{if } y_t = 0,$$

where y_t is the outcome of interest for patient t, with 1 indicating the outcome occurred and 0 indicating it did not occur; p_t is the estimated (prior) risk of having the outcome for patient t; R_0 is the odds ratio under the null hypothesis (often $R_0=1.0$) and R_A is the odds ratio under the alternative hypothesis of a pre-specified clinically important increase (or decrease) in the outcome of interest. The CUSUM signals when $X_t \geq h$ where h is the pre-specified control limit. For the risk-adjustment model in Equation (1): $p_t=1/[1 + \exp(\beta_0 + \beta_1 z)]$.

We assume that there are a discrete number of possible patient types, e.g. Parsonnet scores are integers, so $f(z)$ gives the multinomial probabilities of the various possible values for z and Y is a discrete outcome like 30-day mortality (yes or no). Let us then assume that the true patient mix and risk adjustment are given. Then, when we plan to implement a control chart we would collect some Phase I data and use these data to estimate both the patient mix (based on the observed distribution of Parsonnet scores) and the risk-adjustment model. We will assume here that there are no errors in the measurement of the covariates and that we use the correct covariates in building the risk-adjustment model.

To determine the effect of estimation error, we repeat the following:

(i) Generate some sample Phase I data (z_i, y_i) where $i=1, 2, \ldots, n$, using the (assumed) true patient mix and risk-adjustment model.
(ii) Use the data to estimate the patient mix and the parameters of the risk-adjustment model.
(iii) Use the estimated patient mix and risk-adjustment model to set up a CUSUM chart (i.e. determine the control limit, h) that will yield some desired value for the in-control ARL_0. The appropriate control limit can be determined by repeatedly using the Markov chain approximation [5] with different values of the threshold (control limit) until an appropriate threshold is found (see the Appendix for details).
(iv) Given the selected threshold and the true patient mix and risk-adjustment model, determine the actual in-control ARL_0 and out-of-control ARL_1 (for a realistic and clinically important increase in the true adverse event rate) that would be obtained. ARL_1 is approximated in the same way as ARL_0 but with the adverse event rate increased by the pre-specified clinically important amount, as defined by an odds ratio.

Repeating steps 1–4 many times gives the distribution for the actual ARL_0 and ARL_1. The effect of the estimation

error will be indicated by the variability of the ARLs. To illustrate, let us consider the simplified example from the introduction with three risk groups of patients in a population of around 7000 patients and further assume that a control limit of 3.0 equates to an ARL_0 of 1000. First, we take a random sample of 1000 patients from the population and fit a logistic model to the sample data. Let's say we observe that 55% of our sample patients are at low risk, 30% at medium risk and 15% at high risk. The logistic model fit to these data suggests low-risk patients having 4% risk of infection, medium-risk patients of 10% and high-risk patients a 25% risk of infection. Based on the sample data, we estimate that a control limit of 3.2 is required for an ARL_0 of 1000. Given this control limit of 3.2, the true ARL_0 is actually 1200 for the population of 7000. Therefore, in this simplified example, we can see that the effect of estimation error has resulted in a control limit that is higher than what is required, an ARL_0 higher than what is assumed, and consequently the ARL_1 will be longer than expected, thus it will take longer to detect a clinically important increase in the 30-day mortality rate than expected.

We used both actual data and simulated data to assess the effect of estimation error. The actual data have been used previously [3] and these are data on a cohort of 6994 patients from the UK who underwent coronary bypass graft surgery (CABG) between 1 January 1992 and 31 December 1998. The adverse event of primary interest was 30-day mortality, which occurred in 6.6% of patients. Previous analysis determined that a strong predictor of 30-day mortality was Parsonnet score, which ranged in this cohort from 0 to 71. The fitted risk-adjustment model $\text{logit}(y)=3.63-0.074x$ predicts a 2.5% risk of death at 30 days for patients with Parsonnet=0 and an 84% risk of death at 30 days for patients with Parsonnet=71. The area under the receiver operator characteristic curve (AUC) is a commonly reported measure of goodness of fit for a logistic regression model [6]. The AUC for this model is 0.76, indicating typical discrimination between higher and lower risk patients for a predictive model in the medical literature. For the purposes of this study, we assume that the patient mix of the 6994 patients and the logistic model that describes the relationship between Parsonnet score and 30-day mortality are correct and not subject to error. Based on this cohort, a control limit=2.71 equates to ARL_0 ~1000 patients and an ARL_1 ~110 patients assuming a doubling of the odds of 30-day mortality. In this example, we have assumed that a doubling of the odds of mortality is a clinically meaningful increase; however, we could have assumed a 50% increase of the odds or a tripling of the odds. The magnitude of the pre-specified increase will impact the control chart performance such that the performance is optimal at a true increase that is the same as the pre-specified increase and larger increases will be detected in a more timely fashion than smaller increases.

To determine the effect of estimation error on the variability of ARL, we randomly selected samples with replacement from the 6994 patients in the cohort. The samples were of sizes 760, 1520, 2280, 3040 and 3800 patients resulting in expected number of deaths equal to 50, 100, 150, 200 and 250, respectively (assuming 6.6% mortality). In total, 1000 samples of each sample size were taken. The procedure given above (steps 1–4) was used with specified ARL_0 ~1000 patients. In addition, the variance of the 'true' in-control run length and the 'true' ARL_1 based on a doubling of the odds of 30-day mortality were calculated. This procedure was repeated for ARL_0 ~500 and 1500 using a sample size of 1520 patients to investigate the effect of the specified ARL_0 on estimation error.

In additional analyses, we varied the adverse event rate and the variation in patient mix to determine how these important variables affect estimation error. To investigate the effect of the adverse event rate, we modified the logistic model by changing the size of the intercept term to manipulate the average adverse event rate to 3.5% and also to 12.2%. The model equations specified were $\text{logit}(y)=4.36-0.074x$ and $\text{logit}(y)=2.88-0.074x$, respectively. The variability of the patient mix in the CABG cohort was large (i.e. risks of adverse events in individuals ranged from very small to very large) and hence we lowered variation by restricting sampling to the lower end of the actual Parsonnet distribution (i.e. patients with Parsonnet scores of $\leq$20). This could occur in practice, for example, if only low-risk patients were accepted for the procedure. The model equation for this scenario was $\text{logit}(y)=4.02-0.114x$. For these additional analyses, we bootstrap [7] sampled the Parsonnet scores from the actual CABG cohort and used the modified logistic model to derive patient mortality risks. We then used these estimated risks to generate a random Bernoulli outcome (adverse event=yes or no) with P (adverse event)=risk probability using the RAND ('Bernoulli', risk) function in SAS. In the case where we restricted sampling to patients with Parsonnet scores of $\leq$20 we assumed that the restricted data set was the true patient mix and the fitted logistic model was the correct risk-adjustment model. In a final scenario, we assumed completely homogeneous patients where the adverse event risk was the same for each patient. The model equation in this case was $\text{logit}(y)=2.65$ giving a uniform risk of 6.6% for each patient.

As recommended by Burton *et al.* [8], we wrote a protocol prior to conducting the simulation procedure described above. SAS, version 9.2 for Windows (SAS Institute, Cary, NC) was used for the analysis. PROC SURVEYSELECT was used to select the samples with the starting seeds generated using the system clock. All the estimates obtained from the samples (including control limits and ARLs) are stored in Microsoft Office Excel version 97-2003 worksheets. We chose 1000 samples for each scenario so that we could estimate the variance of the ARL with standard deviation (SD) of <5% of the variance. Each simulation took between 5 and 10 min and hence time was also a factor in deciding how many simulations to run.

Results

Table 1 shows the results of the 1000 samples taken for each of 5 sample sizes. The first row of data is for a sample size

Table 1 Mean (SD) of estimated parameters based on 1000 samples taken for each of the five sample sizes

Sample size (patients)	Number of deaths	Estimated control limit	True ARL_0	True $SDRL_0$	True ARL_1
760	50.1 (7.2)	2.71 (0.11)	1008 (131.5)	1398 (184)	110 (5.7)
1520[a]	100.0 (9.6)	2.72 (0.074)	1011 (89.2)	1402 (125)	110 (3.8)
2280	150.8 (12.2)	2.72 (0.061)	1016 (73.6)	1408 (103)	110 (3.1)
3040	200.8 (13.7)	2.72 (0.053)	1015 (64.4)	1408 (90)	110 (2.7)
3800	249.9 (14.8)	2.72 (0.047)	1014 (56.8)	1407 (79)	110 (2.4)

[a]This is the same reference scenario in each of the five tables.

Table 2 95% confidence intervals (CIs) based on 50 bootstrap samples for SD of true ARL_0 by sample size, event rate and specified ARL_0

Sample size (patients)	Average % event rate	Specified ARL_0	SD (true ARL_0)	95% CI of SD (true ARL_0)
760	6.6	1000	131.5	125.7–137.2
1520[a]	6.6	1000	89.2	86.1–92.3
2280	6.6	1000	73.6	70.3–76.9
3040	6.6	1000	64.4	61.6–67.1
3800	6.6	1000	56.8	54.4–59.3
1520	6.6	500	43.2	41.4–45.0
1520	6.6	1500	129.9	125.7–134.1
1520	3.5	1000	125.9	120.5–131.3
1520	12.2	1000	60.5	58.6–62.4

[a]This is the same reference scenario in each of the five tables.

of 760 patients and specified ARL_0 ~1000. Although a control limit of 2.71 is required for ARL_0 ~1000, due to sampling error the actual threshold values for the 1000 samples ranged between 2.25 and 3.01 (with mean of 2.71 and SD of 0.11). This variation in control limit was sufficient to vary the true ARL_0 from 564 to 1420 (with mean of 1008 and SD of 131.5). The variability in the true ARLs and SD of the true in-control run length ($SDRL_0$) unsurprisingly decrease as the sample size increases. However, on closer inspection, it can be seen that as sample size doubles, the variation halves and hence the SD decreases by a factor of $\sqrt{2}$. If sampling variation is taken into consideration (Table 2), then the observed data are consistent with an exact reciprocal relationship between sample size and variability.

The effect of specifying a different ARL_0 on the estimation error is shown in Table 3 (and Table 2). The data show a relationship between specified ARL_0 and SD of the true ARL_0 where SD doubles as specified ARL_0 doubles. This relationship is also apparent for $SDRL_0$ but not for ARL_1, where the SD of true ARL_1 is only moderately increased for specified ARL_0=1000 compared with specified ARL_0=500 but virtually unchanged for specified ARL_0=1500 compared with specified ARL_0=1000.

Table 4 (and Table 2) show a relationship between adverse event rate and estimation error similar to that observed for sample size. As the event rate (approximately) doubles, the variation in the true ARL and the true SD of run length (approximately) halves. Hence, the SD decreases by a factor of $\sqrt{2}$. However, this relationship does not hold for ARL_1 where there is a reciprocal association between event rate and SD of the true ARL_1. In other words, the effect of event rate on the estimation error effect is more pronounced for ARL_1 compared with ARL_0.

In another analysis, the effect of lower variation in patient mix was investigated. A Parsonnet score of 20 is associated with a 30-day mortality risk of 10% and hence by restricting procedures to patients with Parsonnet scores $\leq$20 we are in effect restricting CABG procedures to patients with a 30-day mortality risk of no $>$10%. This restriction resulted in removing 11% of the total cohort leaving 6211 patients for the analysis with 4.6% of patients dying within 30 days. The results are similar to what would have been expected using all 6994 patients and 69 expected deaths and hence it appears, in this case, a more homogeneous group of patients has had little effect on the effects of estimation error (Table 5). If completely homogeneous patients are assumed, where no risk adjustment is necessary, estimation error is reduced but remains high (Table 5).

Discussion

In this study, we have shown that estimation error can have a substantial effect on risk-adjusted CUSUM chart performance. If we assume that Phase I data include 50 adverse events, a specified ARL_0 of 1000 and a typical risk-adjustment model then the true ARL_0 could be as low as 564 or as high as 1420 based on our 1000 simulations. Therefore, although one false alarm is assumed for every 1000 patients, in fact it could be as high as one in every 564 patients or as low as one in every 1420 patients. In terms of ARL_1, the assumed value is 110 but in fact the true value could be as low as 86 or as high as 125. This difference in ARL_1 between these two extreme scenarios is equivalent to a difference of three additional adverse events, assuming an in-control adverse event rate of 6.6% and a doubling of the odds of adverse event under the alternative hypothesis.

When considering whether to delay Phase II monitoring until a substantial number of events have been included in Phase I data collection (to minimize potential estimation

Table 3 Mean (SD) of estimated parameters based on 1520 patients for each of three specified ARL_0

Specified ARL_0	Number of deaths	Estimated control limit	True ARL_0	True $SDRL_0$	True ARL_1
500	100.4 (9.5)	2.16 (0.066)	504 (43.2)	695 (60)	82 (3.2)
1000[a]	100.0 (9.6)	2.72 (0.074)	1011 (89.2)	1402 (125)	110 (3.8)
1500	100.6 (9.5)	3.07 (0.075)	1526 (129.9)	2123 (182)	128 (3.9)

[a]This is the same reference scenario in each of the five tables.

Table 4 Mean (SD) of estimated parameters based on 1520 patients for each of three adverse event rates

Adverse event rate (%)	Number of deaths	Estimated control limit	True ARL_0	True $SDRL_0$	True ARL_1
3.5	53.8 (7.5)	2.27 (0.10)	989 (125.9)	1367 (175)	148 (8.5)
6.6[a]	100.0 (9.6)	2.72 (0.074)	1011 (89.2)	1402 (125)	110 (3.8)
12.2	185.2 (12.9)	3.15 (0.054)	1001 (60.5)	1392 (85)	81 (1.7)

[a]This is the same reference scenario in each of the five tables.

Table 5 Mean (SD) of estimated parameters based on 1520 patients for each of three levels of patient variability

Patient variability	Number of deaths	Estimated control limit	True ARL_0	True $SDRL_0$	True ARL_1
High[a]	100.0 (9.6)	2.72 (0.074)	1011 (89.2)	1402 (125)	110 (3.8)
Low	69.5 (7.9)	2.54 (0.085)	1003 (103.4)	1390 (144)	123 (5.3)
None	100.4 (9.5)	2.85 (0.069)	999 (79.7)	1389 (112)	98 (3.0)

[a]This is the same reference scenario in each of the five tables.

error), it might be better to avoid delay and move to an initial Phase II stage where model parameters continue to be updated but patient monitoring is also implemented prospectively. However, care is needed to ensure that model parameters are updated using only in-control observations and avoiding observations from periods where the system is out of control. In addition, the effect of estimation error when designing a control chart could be taken into account. This could be done by generating a number of bootstrap samples of the available phase I data and then, for each sample, determining the control limit needed to obtain an ARL0 of a pre-specified level. To determine the ultimate control limit to implement, a value that maintains an ARL_0 of at least the pre-specified level in (say) 95% of the samples could be chosen. For example, in our scenario of 100 adverse events and a specified ARL_0 of 1000, a control limit of 2.84 maintains an ARL_0 of at least 1000 in 95% of the samples. This contrasts with an average control limit of 2.71 if we simply aim for an ARL_0 of 1000. However, this is a conservative strategy with the trade-off that the control chart will be slower to detect a clinically important deterioration in the adverse event rate.

Variables thought to potentially influence estimation error considered in this research included Phase I sample size, number of adverse events, specified ARL_0 and patient variability. The two most important variables were unsurprisingly the number of events and the specified ARL_0. In terms of the number of adverse events, as the number of events doubles, the variance of the true ARL_0 halves. However, interestingly the effect on true ARL_1 is even more pronounced where it is the SD that halves. In the case of specified ARL_0, the effect of estimation error decreases as the specified ARL_0 decreases. This relationship was strong for true ARL_0 where the SD halved when the specified ARL_0 was halved; however, the relationship was weak for true ARL_1 where the SD only decreased slightly when the specified ARL_0 was reduced by a factor of 3. A third variable, patient risk variability, appeared to have little effect on the estimation error. In the case of homogeneous patients where adverse event risk was assumed to be constant at 6.6%, the estimated level of estimation error: SD $(ARL_0)=79.7$ was less than the equivalent risk-adjusted scenario where SD $(ARL_0)=89.2$ but only by around 10%. This result suggests that it is the uncertainty in the overall adverse event rate that is the main component of estimation error. The uncertainties of the patient mix and risk-adjustment model, whilst not ignorable, only appear to account for a modest proportion of the total estimation error.

There are a number of important limitations of this study. First of all, we only used one example data set; therefore, our results may not generalize to all relevant situations where risk-adjusted control charts are implemented. In addition, we

only considered risk-adjusted CUSUM charts where risk adjustment was done by the method of Steiner *et al.* [3]. However, we did use a more general set of simulated patients in the analyses that investigated the effects of event rate and patient variability on estimation error. We also varied a number of important variables such as sample size and specified ARL_0 to make our results more generalizable. Another limitation is that we used an approximate method for determining the ARLs. The Markov chain method is associated with some error due to the state space being divided into a discrete number of states. However, we minimized this error by including a large number of states in the transition matrices: between 1000 and 1500 in most cases. With these large matrices, each simulation took around 5–10 min to run. A further limitation was our use of simulations to investigate the relationship between estimation error and control chart performance. Therefore, our results are not exact; however, we included 1000 simulations for each scenario investigated and, therefore, the 95% confidence intervals for estimated SDs of the true ARLs were relatively narrow. These confidence intervals were sufficiently tight to allow us to determine the exact relationships between important variables such as number of events and the magnitude of estimation error. A final limitation is that we have assumed that the risk-adjustment model is correct and not subject to confounding bias. If a risk-adjustment model is not correct, then the results of this study would not be applicable.

References

1. Montgomery DC. *Introduction to Statistical Quality Control*, 6th edn. New York: John Wiley & Sons, 2008, 198–9.
2. Jensen WA, Jones-Farmer LA, Champ CW, Woodall WH. Effects of parameter estimation on control chart properties: a literature review. *J Qual Technol* 2006;**38**:349–64.
3. Steiner SH, Cook RJ, Farewell VT, Treasure T. Monitoring surgical performance using risk-adjusted cumulative sum charts. *Biostatistics* 2000;**1**:441–52.
4. Parsonnet V, Dean D, Bernstein AD. A method of uniform stratification of risk for evaluating the results of surgery in acquired adult heart disease. *Circulation* 1989;**79**:I3–12.
5. Brook D, Evans DA. An approach to the probability distribution of Cusum run length. *Biometrika* 1972;**59**:539–49.
6. Zou KH, O'Malley AJ, Mauri L. Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. *Circulation* 2007;**115**:654–7.
7. Efron B, Tibshirani RJ. *An Introduction to the Bootstrap*. Boca Raton, FL: CRC Press, 1994.
8. Burton A, Altman DG, Royston P *et al.* The design of simulation studies in medical statistics. *Stat Med* 2006;**25**:4279–92.
9. Steiner SH. Grouped data exponentially weighted moving average control charts. *Appl Stat* 1998;**47**:203–16.

Appendix

Markov chain approximation for estimating average and variance of run lengths

Average and variance of run lengths were estimated using a Markov chain approximation developed by Brook and Evans [5]. This methodology is approximate because the state space is divided into a discrete number of states with the last state being the out-of-control condition. By using a large number of states as we have done, the approximation is very accurate. CUSUM control charts are based on a Markov process where the current value of the control chart depends only on the previous value. The transition from the current value to the next value can be described completely by the possible future values and the probabilities that these values will occur. Therefore, the process can be represented by a transition matrix. Each row of the matrix represents the current value and the columns contain the probabilities of moving to each and every one of the possible values of the control chart. In cases where the transition from the current value to a particular future value is impossible, the transition probability is zero. The final row and column of the transition matrix represents the absorbing state which occurs when a control chart value is obtained that is greater than the control limit. Hence, the last row and column represents a range of possible control chart values bounded below by the control limit. The size of the transition matrix is therefore $(n+1)$ by $(n+1)$, where n represents the n possible values that the control chart can take prior to exceeding the control limit. If the last row and column is deleted from the transition matrix, then ARL and variance of the run length are given by

$$E(\gamma) = \sum_{t=1}^{\infty} t \Pr(\gamma = t) \sum_{t=1}^{\infty} R'1 = (I - R)^{-1}\mathbf{1}.$$
$$\mathrm{var}(\gamma) = 2R(I - R)^{-2}\mathbf{1}.$$

where γ is the run length; t the time; R the transition matrix after last row and column have been deleted; $\mathbf{1}$ the column vector of ones and I the identity matrix (see the Appendix of Steiner [9] for more details).