



Universiteit  
Leiden  
The Netherlands

## **CGR-CUSUM: a continuous time generalized rapid response cumulative sum chart**

Gomon, D.; Putter, H.; Nelissen, R.G.H.H.; Pas, S.L. van der

### **Citation**

Gomon, D., Putter, H., Nelissen, R. G. H. H., & Pas, S. L. van der. (2022). CGR-CUSUM: a continuous time generalized rapid response cumulative sum chart. *Biostatistics*. doi:10.1093/biostatistics/kxac041

Version: Publisher's Version  
License: [Creative Commons CC BY 4.0 license](https://creativecommons.org/licenses/by/4.0/)  
Downloaded from: <https://hdl.handle.net/1887/3479641>

**Note:** To cite this publication please use the final published version (if applicable).



# CGR-CUSUM: a continuous time generalized rapid response cumulative sum chart

DANIEL GOMON\*

*Department of Statistics, Mathematical Institute, Leiden University, Niels Bohrweg 1, 2333CA Leiden, The Netherlands.*

d.gomon@math.leidenuniv.nl

HEIN PUTTER

*Department of Biomedical Data Sciences, Leiden University Medical Centre, Einthovenweg 20, 2333ZC Leiden, The Netherlands*

ROB G. H. H. NELISSEN

*Department of Orthopaedic Surgery, Leiden University Medical Centre, Leiden, Albinusdreef 2, 2333 ZA Leiden, The Netherlands*

STÉPHANIE VAN DER PAS

*Department of Epidemiology and Data Science, Amsterdam UMC, Vrije Universiteit Amsterdam, De Boelelaan 1089A, 1081HV Amsterdam, The Netherlands*

## SUMMARY

Rapidly detecting problems in the quality of care is of utmost importance for the well-being of patients. Without proper inspection schemes, such problems can go undetected for years. Cumulative sum (CUSUM) charts have proven to be useful for quality control, yet available methodology for survival outcomes is limited. The few available continuous time inspection charts usually require the researcher to specify an expected increase in the failure rate in advance, thereby requiring prior knowledge about the problem at hand. Misspecifying parameters can lead to false positive alerts and large detection delays. To solve this problem, we take a more general approach to derive the new Continuous time Generalized Rapid response CUSUM (CGR-CUSUM) chart. We find an expression for the approximate average run length (average time to detection) and illustrate the possible gain in detection speed by using the CGR-CUSUM over other commonly used monitoring schemes on a real-life data set from the Dutch Arthroplasty Register as well as in simulation studies. Besides the inspection of medical procedures, the CGR-CUSUM can also be used for other real-time inspection schemes such as industrial production lines and quality control of services.

**Keywords:** Benchmarking; Continuous time; Control charts; CUSUM; Generalized likelihood ratio; Quality of care; Survival analysis.

\*To whom correspondence should be addressed.

## 1. INTRODUCTION

Rapid detection of deterioration in the quality of care can spare patients unnecessary health burdens. There are currently many inspection schemes that can be used to monitor the quality of care, such as funnel plots (Spiegelhalter, 2005) and a variety of Cumulative sum (CUSUM) charts (Steiner *and others*, 2000; Biswas and Kalbfleisch, 2008). A particularly attractive property of CUSUM charts is that they can be used to sequentially check for a decrease in the quality of a process. Ideally, the inspection scheme is also tailored to the outcome type. In this article, we are interested in inspecting survival outcomes, where every individual can experience a failure at any time after their entry into the study. As an example, the Dutch Arthroplasty Register (LROI) is interested in simultaneously monitoring the quality of orthopedic care at multiple hospitals performing total hip replacement surgery by considering the information provided by the time of implant failure as soon as it occurs, as well as the information provided by patients not experiencing implant failures. To facilitate such real-time inspection, Biswas and Kalbfleisch (2008) developed a CUSUM chart for survival outcomes, followed by Sego *and others* (2009) and Begun *and others* (2019). Each of these charts uses different assumptions in the CUSUM model and is therefore applicable in different scenarios. One similarity is that all of them require the researcher to specify an expected increase in the future rate of failure. When this quantity is chosen incorrectly, the charts may experience delays in detection and produce false negative signals.

Our main goal in this article is to develop a method that no longer requires the researcher to specify many parameters in advance, thereby requiring less prior information for inspection and leading to faster detection times in practical applications. For this reason, we devise a generalization of the CUSUM chart by Biswas and Kalbfleisch (2008), which we call the Continuous time Generalized Rapid response CUSUM (CGR-CUSUM). Biswas and Kalbfleisch (2008) chose to only consider the information provided by patients until 1 year after their procedure. In contrast, the CGR-CUSUM is constructed using all available information on any patient at all times. A consequence of these changes is that generally our chart leads to quicker detection of underperforming hospitals, thereby contributing to the improvement of the quality of care.

Other methods for the continuous time inspection of the quality of care include the uEWMA chart for survival time data by Steiner and Jones (2009) and the STRAND chart by Grigg (2018). Grigg (2018) briefly discusses the differences among the BK-CUSUM, uEWMA, and STRAND charts and concludes that the uEWMA and STRAND charts are particularly suitable for quick detection when failures are clustered. In contrast, the BK-CUSUM and the CGR-CUSUM are designed to detect increased failure rates without a specific mechanism for clusters.

We derive an approximation for the average run length (average time to detection) of the CGR-CUSUM, by means of considering a simplification of the CGR-CUSUM called the Continuous time Generalised Initial response CUSUM (CGI-CUSUM). Additionally, we consider an adjusted Biswas and Kalbfleisch (2008) CUSUM procedure which uses the information of all patients at all times, which we call the BK-CUSUM for convenience. Similarly, we present an approximation to the average run length of the BK-CUSUM and compare this approximation with the approximation found for the CGR-CUSUM. This comparison demonstrates how incorrect prior information can significantly increase the detection times of the BK-CUSUM procedure, which then also carries over to the Biswas and Kalbfleisch (2008) CUSUM.

The new CGR-CUSUM chart can be a very useful tool in practical applications where the future expected rate of failure is not known in advance or likely to vary over the time of the study. As this occurs often in medical applications, the CGR-CUSUM chart can significantly improve the quality of care worldwide by inspecting current procedures. In contrast to the multi-chart CUSUM scheme of Han and Tsung (2007), where the possible increase in failure rate is considered over a finite probable domain, the CGR-CUSUM only requires the construction of one chart and the increase in failure rate can also be limited to a fixed domain. On top of this, the CGR-CUSUM is not limited to medical applications.

The chart can be used to inspect any procedures involving “survival” outcomes, such as production lines and customer satisfaction inspection.

In Section 2 of this article, the relevant quantities, notation, the CGR-CUSUM and BK-CUSUM are introduced. An approximate average run length is derived for both procedures. In Section 3, all methods are applied to a data set from the LROI. In Section 4, a simulation study is performed to compare the average run lengths of aforementioned procedures under restrictions on their null (hypothesis) average run length. Additionally, a simulation study is performed using the data from this register where the type I error of the charts over time is restricted under the null rate. The article concludes with a discussion and recommendations for practice.

## 2. METHODS

### 2.1. Model and data

Following Biswas and Kalbfleisch (2008), consider a hospital with subjects  $i = 1, 2, \dots$  arriving (entering the study) according to a Poisson process with rate  $\psi$ . Let  $S_i$  denote the time of the entry of subject  $i$  into the study, relative to the starting time  $t = 0$ . Denote by  $X_i$  the time from entry until failure, such that  $T_i = S_i + X_i$  is the chronological time of failure. Consider only right-censored observations, and let  $R_i$  denote the chronological time of right-censoring of observation  $i$ . Let the  $p$ -vector  $\mathbf{Z}_i$  denote the relevant covariates of subject  $i$ . Assume that there is a known null distribution for the subject-specific time to failure, denoted by the hazard rate  $h_i(x)$ . We make use of the Cox proportional hazards model to incorporate the covariates, such that  $h_i(x) = h_0(x) \exp(\mathbf{Z}_i^\top \boldsymbol{\beta})$  with regression coefficients  $\boldsymbol{\beta}$  and known baseline hazard rate  $h_0$ . Let  $Y_i(t) = \mathbb{1}\{S_i \leq t \leq \min\{T_i, R_i\}\}$  be an indicator whether subject  $i$  is active at time  $t$ . Define  $\tilde{N}_i(t) = \mathbb{1}\{T_i \leq t\}$  and subsequently define  $N_i(t) = \int_0^t Y_i(u) d\tilde{N}_i(u)$  for  $t > 0$  as the counting process for an observed failure of subject  $i$ . Define  $N(t) = \sum_{i \geq 1} N_i(t)$  as the counting process for the total number of failures observed at the hospital. Define the cumulative intensity of subject  $i$  as  $\Lambda_i(t) = \int_0^t \lambda_i(u) du$  with  $\lambda_i(u) = Y_i(u) \cdot h_i(u)$ . Let the superscript  $\theta$  indicate an increase in the hazard rate such that  $\Lambda_i^\theta(t) = \Lambda_i(t) \cdot \exp(\theta)$  and  $h_i^\theta(t) = h_i(t) \cdot \exp(\theta)$  and  $F_i^\theta$  the associated cumulative distribution function. We call  $\exp(\theta)$  the hazard ratio and say that the process is *in control* when  $\theta = 0$  and *out of control* when  $\theta > 0$ . Define  $\Lambda(t) = \sum_{i \geq 1} \Lambda_i(t)$  as the total cumulative intensity at the hospital at time  $t$ . For aforementioned counting processes, define  $dN(t) = N(t + dt) - N(t)$ , with  $dt$  an infinitesimally small quantity. It follows that:

$$\mathbb{P}(dN_i(t) = 1 | T_i \geq t, S_i, Z_i) = \begin{cases} e^\theta h_i(t - S_i) dt, & \text{if } 0 \leq t - S_i, \\ 0, & \text{else.} \end{cases} \quad (2.1)$$

We denote the right-hand side of (2.1) by  $d\Lambda_i^\theta(t)$ . Finally, let  $\Delta N_i(t) = N_i(t) - N_i(t-)$  be the increment of  $N_i$  at time  $t$  with  $t-$  the time “just before” time  $t$ .

### 2.2. Continuous time Generalized Rapid response CUSUM

The CUSUM procedure developed by Biswas and Kalbfleisch (2008) can be used to test whether the hazard rate at a hospital has increased from  $\Lambda_i$  to  $\Lambda_i^\theta$  for some fixed and known  $\theta > 0$ , at some unknown time after the start of the study. This procedure is very useful when there is some prior knowledge about the true hazard ratio  $e^\theta$ , but may lead to delays in detection when this is not the case or when the rate of failure is variable. For this reason, we will consider a more general test, where the expected hazard ratio no longer needs to be specified in advance, much like the GLR Statistic in Siegmund and Venkatraman (1995) is a generalization of the original CUSUM procedure of Page (1954).

To achieve this, we test the null hypothesis of no change against the alternative of a sudden change in hazard rate at some unknown time  $s > 0$ , affecting all subjects at risk at time  $s$  and thereafter:

$$H_0 : X_i \sim \Lambda_i(t), i = 1, 2, \dots \quad H_1 : \begin{cases} X_i \sim \Lambda_i(t) | t < s, i = 1, 2, \dots \\ X_i \sim \Lambda_i^\theta(t) | t \geq s, i = 1, 2, \dots \end{cases} \quad (2.2)$$

with  $\theta > 0$ . Let us find the likelihood ratio for a test of  $\theta = 0$  against  $\theta = \theta_1$  with  $\theta_1 > 0$  an unknown constant. The likelihood for the aggregated counting process  $N(t)$  at study time  $t$  with  $n$  subjects is then given by  $\prod_{i=1}^n \left( \prod_{0 \leq u \leq t} \lambda_i^\theta(u)^{\Delta N_i(u)} \exp(-\Lambda_i^\theta(t)) \right)$  (see [Aalen and others, 2008](#), Section 5.1). Note that  $\Delta N_i(u)$  is non-zero only at the time of failure of subject  $i$ , where it is equal to one. This yields a likelihood ratio statistic at time  $t$  of:

$$\begin{aligned} \text{LR}(t) &= \sup_{\theta \geq 0} \frac{\prod_{i=1}^n \left( \prod_{0 \leq u \leq t} \exp(\theta) \lambda_i(u)^{\Delta N_i(u)} \right) \exp(-\exp(\theta) \Lambda_i(t))}{\prod_{i=1}^n \left( \prod_{0 \leq u \leq t} \lambda_i(u)^{\Delta N_i(u)} \right) \exp(-\Lambda_i(t))} \\ &= \prod_{i=1}^n \frac{\left( \exp(\hat{\theta}(t)) \right)^{N_i(t)} \exp(-\exp(\hat{\theta}(t)) \Lambda_i(t))}{\exp(-\Lambda_i(t))}, \end{aligned}$$

where  $\hat{\theta}(t)$  is the maximum likelihood estimate of  $\theta$  at time  $t$ . This maximum likelihood estimator  $\hat{\theta}(t)$  can be determined by maximizing the likelihood at a hospital where patients are failing with cumulative intensity  $e^\theta \Lambda_i(t)$  up to time  $t$  over  $\theta$  and is given by:

$$\hat{\theta}(t) = \max \left( 0, \log \left( \frac{N(t)}{\Lambda(t)} \right) \right). \quad (2.3)$$

The logarithm of the LR statistic is then given by:

$$U(t) := \log(\text{LR}(t)) = \hat{\theta}(t)N(t) - \left( \exp(\hat{\theta}(t)) - 1 \right) \Lambda(t).$$

Note that this quantity will increase when a failure is observed, and drift downwards at all other times. A preliminary chart is then given by:

$$G(t) = \max_{s: 0 \leq s \leq t} \left\{ \hat{\theta}(s, t)N(s, t) - \left( \exp(\hat{\theta}(s, t)) - 1 \right) \Lambda(s, t) \right\}, \quad (2.4)$$

where  $s$  indicates that the quantity is determined using the information provided by all active patients in the time frame  $(s, t)$ :

$$\begin{aligned} N(s, t) &= N(t) - N(s) \\ \Lambda(s, t) &= \Lambda(t) - \Lambda(s) \end{aligned} \quad \text{and} \quad \hat{\theta}(s, t) = \max \left( 0, \log \left( \frac{N(s, t)}{\Lambda(s, t)} \right) \right). \quad (2.5)$$

In contrast to the method developed by [Biswas and Kalbfleisch \(2008\)](#), it is not possible to determine this chart recursively as the maximum likelihood estimator needs to be determined over multiple time frames. This makes the chart very computationally expensive. We therefore consider simpler hypotheses:

$$H_0 : X_i \sim \Lambda_i, i = 1, 2, \dots \quad H_1 : \begin{cases} X_i \sim \Lambda_i, i = 1, 2, \dots, \nu - 1 \\ X_i \sim \Lambda_i^\theta, i = \nu, \nu + 1, \dots \end{cases} \quad (2.6)$$

with  $\theta > 0$  and  $\nu \geq 1$  both unknown in advance. We then test the null hypothesis of no change against the alternative that the rate of failure at the hospital has increased to  $e^\theta \Lambda_i$ , starting from some subject  $\nu \geq 1$ . These hypotheses make sense in a medical context, where the hazard rate is likely to depend on the entry time of the patient.

DEFINITION 1 The continuous time generalized rapid response CUSUM (CGR-CUSUM) chart is given by:

$$\text{CGR}(t) = \max_{\nu \geq 1} \left\{ \hat{\theta}_{\geq \nu}(t) N_{\geq \nu}(t) - \left( \exp\left(\hat{\theta}_{\geq \nu}(t)\right) - 1 \right) \Lambda_{\geq \nu}(t) \right\} \quad (2.7)$$

with (subjects sorted according to chronological arrival time):

$$\begin{aligned} N_{\geq \nu}(t) &= \sum_{i \geq \nu} N_i(t) & \text{and} & \hat{\theta}_{\geq \nu}(t) = \max \left( 0, \log \left( \frac{N_{\geq \nu}(t)}{\Lambda_{\geq \nu}(t)} \right) \right). \end{aligned} \quad (2.8)$$

In the CGR-CUSUM patients prior to the  $\nu$ th patient no longer contribute to the chart at all, whereas in  $G(t)$  all patients active after time  $t - s$  still contribute to the value of the chart. This difference is highlighted in [Figure S1](#) of the [Supplementary material](#) available at *Biostatistics* online. To employ a testing procedure, we construct the chart  $\text{CGR}(t)$  at every relevant time point  $t$  and reject the null hypothesis (producing a signal) as soon as  $\text{CGR}(t) \geq h$  for some  $h > 0$ . This constant  $h$  is called the *control limit* and can be chosen in accordance with some desired property of the procedure such as the average run length of the chart defined below.

DEFINITION 2 Denote by  $\tau_h = \inf\{t > 0 : \text{CGR}(t) \geq h\}$  the time it takes for a CGR-CUSUM to produce a signal. The average run length (ARL) is then defined as  $\mathbb{E}[\tau_h]$ . We refer to the *in control* average run length as the expected time to detection when  $\exp(\theta) = 1$  and *out of control* average run length when  $\exp(\theta) > 1$ .

### 2.3. An approximation to the ARL

In this section, we will derive an upper bound for the average run length of the CGR-CUSUM in the out-of-control case. The maximization term in (2.7) poses a significant challenge in approximating the ARL. It turns out that we can derive a bound on the ARL through comparison with a simpler version of the CGR chart. For this reason, we consider the Continuous time Generalised Initial response (CGI) CUSUM chart. This chart can be used to test the hypotheses of an initial change in the rate of failure:

$$H_0 : X_i \sim \Lambda_i, i = 1, 2, \dots \quad H_1 : X_i \sim \Lambda_i^\theta, i = 1, 2, \dots$$

DEFINITION 3 The Continuous time Generalized Initial response CUSUM (CGI-CUSUM) with  $\hat{\theta}(t)$  as in (2.3) is given by:

$$\text{CGI}(t) = \hat{\theta}(t) N(t) - \left( \exp(\hat{\theta}(t)) - 1 \right) \Lambda(t).$$

Note how the CGI chart is simply the CGR chart without the maximization term. The CGI-CUSUM is not a chart which should be used in practice as it cannot be used to sequentially detect a changepoint in the process, but instead it is merely a tool for theory. Due to its simpler expression, it is possible to determine the asymptotic distribution of the chart under some assumptions. One of the key assumptions

is that subjects arrive according to a Poisson process with rate  $\psi$ , allowing us to equate the number of patients to time by  $n \approx \psi \cdot t$ .

**THEOREM 2.1** Suppose that subjects arrive according to a Poisson process with rate  $\psi$  under suitable regularity conditions. Then, for  $\theta > 0$ :

$$\sqrt{t}(\text{CGI}(t) - (\theta + \exp(-\theta) - 1)I(\theta, t)) \xrightarrow{d} \mathcal{N}(0, t\theta^2 I(\theta, t)),$$

and when  $\theta = 0$  (using the shape  $k$ /scale  $b$  parametrization):

$$t \cdot \text{CGI}(t) \xrightarrow{d} \Gamma\left(k = \frac{1}{2}, b = t\right)$$

where  $I(\theta, t) = \psi \int_0^t \mathbb{E}_{Z_i} [F_i^\theta(s)] ds$  is the Fisher information in all observations at time  $t$ .

The proof of this theorem, the required regularity conditions as well as the derivation of the Fisher information can be found in the [Supplementary materials Sections 2, 3, and 4](#). The usefulness of this result depends on the availability of an expression for  $I(\theta, t)$ . A discussion on how to calculate the Fisher information, as well as some examples for the PVF family of distributions can be found in [Section S7](#) of the [Supplementary material](#) available at *Biostatistics* online. We determine an approximate (asymptotic) average run length for the CGI chart by equating the expected value of the asymptotic distribution to the control limit  $h$ .

**LEMMA 2.2** We find an approximate average run length  $\widehat{\text{ARL}}_{\text{CGI}}(\theta, h)$  for the CGI-CUSUM when  $\exp(\theta) > 1$  by solving the following equation for  $t$ :

$$(\theta + \exp(-\theta) - 1)I(\theta, t) = h. \quad (2.9)$$

For  $\exp(\theta) = 1$ , this method yields no approximation to the ARL, and it is therefore not possible to determine theoretical control limits which restrict the in control ARL. It is possible to approximate the value of the in control average run length by means of Monte Carlo simulation when it is of interest. Note that due to the convergence requirement, this approximate ARL will not yield good approximations for small values of the control limit  $h$ . The theoretical out-of-control ARL will be evaluated by means of simulation in [Section 4.1](#).

Note that the CGR-CUSUM is simply a CGI-CUSUM maximized over the last  $n - \nu$  patients. As a result, the CGR-CUSUM is always larger or equal than the CGI-CUSUM. This property allows us to compare the average run lengths of the CGR- and CGI-CUSUM charts.

**REMARK 2.3** Suppose that subjects are failing with an increased hazard rate  $\Lambda_i \exp(\theta)$  from the beginning of the study. Then the average run lengths of the charts can be compared as follows:

$$\text{ARL}_{\text{CGR}}(\theta, h) \leq \text{ARL}_{\text{CGI}}(\theta, h).$$

In most practical applications, an upper bound is sufficient as the interest lies in restricting the run time of the chart from above when the failure rate is higher than expected.

Due to the found upper bound, we can now determine the CGI chart on out-of-control samples in simulation studies to obtain information on the ARL of the CGR chart for comparable samples. This



negates the need to construct the CGR chart when approximating the ARL, saving a lot of computation time. Another way to reduce the computation time of the CGR- and CGI-CUSUM charts is given in the following corollary.

REMARK 2.4 The value of the CGR-CUSUM and CGI-CUSUM can only increase at a time point when a failure is observed. As a consequence, for detection purposes it is sufficient to only determine the value of the charts at the times of failure.

#### 2.4. The Biswas and Kalbfleisch (2008) CUSUM and CGR-CUSUM

By a priori fixing a value  $\theta_1 > 0$  for  $\theta$  in the chart  $G(t)$  (see (2.4)) we would recover the CUSUM procedure developed by Biswas and Kalbfleisch (2008). The biggest advantage of the CGR-CUSUM over the Biswas and Kalbfleisch (2008) CUSUM is that we no longer need to specify this expected hazard ratio, allowing for a more general test requiring less prior knowledge. Besides this, the maximum likelihood estimator allows for updating the parameter to the most recent failure rates. In contrast, the maximum likelihood estimator needs time to converge to the true value, possibly causing delays in detection when compared to the Biswas and Kalbfleisch (2008) CUSUM with correctly specified  $\theta_1$ .

Biswas and Kalbfleisch (2008) note that 1-year postprocedure survival outcomes are often employed for medical inspection schemes and decide to consider subjects as active only for  $C = 1$  year after the procedure. This limitation allows them to derive a closed-form approximation to the average run length of the chart. We decide not to disregard the information provided by patients 1-year postprocedure. The value of the chart is then based on more complete information, possibly leading to quicker detection times. With this approach, determining an expected run length shorter than  $C = 1$  year is possible, in contrast to Biswas and Kalbfleisch (2008). Our new approach then also leads towards an approximate ARL for the Biswas and Kalbfleisch (2008) CUSUM procedure with the  $C = 1$  limitation relaxed. Further on in this article, we will only consider the Biswas and Kalbfleisch (2008) CUSUM procedure with the  $C = 1$  limitation relaxed, as it is more similar to our CGR chart. We call this chart the BK-CUSUM chart.

DEFINITION 4 The BK-CUSUM is given by:

$$BK(t) = \max_{s: 0 \leq s \leq t} \{\theta_1 N(s, t) - (\exp(\theta_1) - 1) \Lambda(s, t)\},$$

with notation as in (2.5) where  $\exp(\theta_1)$  is the expected hazard ratio chosen in advance.

Taking a similar approach to Section 2.3, it is possible to determine an approximate average run length for the BK-CUSUM procedure.

COROLLARY 1 Suppose  $\widehat{\theta}_1$  is chosen such that  $\exp(\widehat{\theta}_1)/\exp(\theta) < \widehat{\theta}_1 + \exp(-\theta)$ . We find an approximate average run length  $\widehat{ARL}_{BK}(\theta, h)$  by solving the following equation for  $t$ :

$$\left( \widehat{\theta}_1 + \exp(-\theta) - \frac{\exp(\widehat{\theta}_1)}{\exp(\theta)} \right) I(\theta, t) = h. \quad (2.10)$$

The proof can be found in Section S5 of the Supplementary material available at *Biostatistics* online.

Due to the restriction on  $\theta_1$  it is not always possible to use this expression for the approximate ARL. As  $I(\theta, t)$  is non-negative for every  $t \geq 0$ , the approximate ARL for the CGR and the BK-CUSUM can be compared. It can easily be seen that when  $\theta_1 \neq \theta$ , the left side of (2.9) is guaranteed to be larger than the



left side of (2.10) for  $t > 0$ . This means that when the expected hazard ratio  $\exp(\theta_1)$  is misspecified, the approximate ARL of the CGI chart will be smaller than that of the BK-CUSUM chart therefore yielding faster out-of-control detection speeds.

The difference between the CGR-CUSUM and BK-CUSUM lies in the hypotheses used for constructing the chart, where the CGR-CUSUM is used to detect a change in hazard rate for all patients entering after some patient entry time and the BK-CUSUM to detect a spontaneous change in hazard rates for all patients at risk after some chronological time. This difference is shown visually in Figure S1 of the Supplementary material available at *Biostatistics* online.

### 3. APPLICATION TO LROI

We demonstrate the possible gain in detection speed when using the CGR-CUSUM over the BK-CUSUM by applying both methods on a hip replacement data set from the LROI. The LROI is the Dutch national registry of all orthopedic implants (e.g., hip, elbow, wrist, ankle, knee, shoulder, finger, and thumb), with a reported completeness of more than 95% for registered hip and knee surgical procedures (van Steenberg and others, 2015; Dutch Arthroplasty Register (LROI), 2020).

#### 3.1. The data set

The data used for the analysis consists of information on total hip replacement surgeries at 97 hospitals across the Netherlands from 01/01/2014 up until 01/01/2020 and was received under agreement LROI 2020 – 053. Available variables are the dates of all primary procedures, time until failure of the prosthesis (our main interest), and/or death of the patient as well as multiple characteristics of each patient which can be found in Table S2 of the Supplementary material available at *Biostatistics* online. Three characteristics of patients had more than 0.5% of missing values, which were BMI (1.8%), Smoking indicator (4.5%) and Charnley Score (5.3%). Using the R package mice (van Buuren and Groothuis-Oudshoorn, 2011), we imputed missing values to obtain a complete data set.

#### 3.2. Baseline: yearly funnel plot

The current method employed by the LROI for comparing implant surgery performance between hospitals is a yearly risk-adjusted funnel plot over all available data of the recent 3–6 years. The funnel plot uses 1-year postsurgery failure as binary outcome, therefore not allowing for continuous inspection of the quality of care. van Schie and others (2020) have used the funnel plot as the “golden standard” for the LROI, indicating which hospitals had problems in their quality of care. As we have no information on the true failure rate and problems at the hospitals in question, we will compare detection times with the funnel plot as well.

#### 3.3. The baseline hazard

In any practical application, the determination of the baseline is of great importance when considering the BK- and CGR-CUSUM charts as this greatly influences the detection speed and false detection rate. In both cases, the baseline is completely fixed by a null hazard rate and the corresponding Cox regression coefficients. In good scientific practice, these quantities should be determined using an in control data set, where failures are known to be happening at an acceptable rate. In reality, it is often difficult to obtain such a set for many different reasons. Because of this, we determine the null hazard rate and Cox coefficients using the whole data set as training set. This implies that the average national failure rate over all hospitals is up to the desired standard. The same is done for the funnel plot.

The yearly funnel plot requires a yearly determination of the baseline. To make a fair comparison between the funnel plot and CUSUM methods, we therefore determine the baseline hazard rate, failure

proportion, and risk-adjustment coefficients using the whole data set restricted to the first 3, 4, 5, and 6 years of information for both methods. This was achieved by using the Cox proportional hazards function in the R package *survival* developed by Therneau (2020). This way, all methods use the same information for the construction of the charts and thereby all use the same “standard of care.” We start with determining parameters at the 3-year margin to have a sufficient amount of information for the determination of the null parameters.

### 3.4. Determining control limits

We determine control limits for the risk-adjusted BK- and CGR-CUSUM charts by restricting the simulated probability of a type I error to 0.05 over a period of 6 years. The procedure is described in Section S6 of the [Supplementary material](#) available at *Biostatistics* online. Due to the extremely low failure rate in the data, we chose to restrict  $e^{\hat{\theta}^{(t)}} \leq 6$ . This comes down to believing the hazard rate at a hospital cannot be more than six times the baseline. Without this limitation, the CGR-CUSUM made very large jumps for patients experiencing near instant failure (first or second day after surgery), almost always leading to detection. The determined control limits and a summary of detection times with respect to the hospitals detected by the funnel plot in the first 3 years can be found in Table 2. The continuous time procedures have their detection time rounded upwards to the closest month, to show what detection times are realistic when constructing the charts monthly. We also include the detection times achieved by the monthly Bernoulli CUSUM procedure suggested by van Schie and others (2020). They chose to take a control limit of  $h = 3.5$  in correspondence to other literature. Exact detection times for all detected hospitals can be found in Table S1 of the [Supplementary material](#) available at *Biostatistics* online.

### 3.5. Result: average detection delays

As the Bernoulli CUSUM and funnel plot both use one year post implant failure as outcome and the same risk-adjustment model, they can both detect exactly the same hospitals at the end of the third year. This is different for the continuous time CUSUM charts: both the BK- and CGR-CUSUM do not detect hospitals 5 and 19 and yielded one and two “false” detections, respectively. Overall, the continuous time CUSUM procedures yield (much) faster detection times, but also signal very different hospitals than the discrete time methods, especially after the 4-year mark. There are multiple reasons for this. We will explain some of the possible mechanisms which cause a mismatch in detections between the methods by means of some examples in Figure 1. A general observation is that the Bernoulli CUSUM has a 1-year delay compared to the continuous time charts. In Figure 1(a), we can see that hospital 5 was signaled by the funnel plot and Bernoulli CUSUM but not by the BK- and CGR-CUSUM charts. Whereas the continuous time charts show a downward motion after a period of multiple consecutive failures, the Bernoulli CUSUM does not. This is most likely due to the fact that the Bernoulli CUSUM and funnel plot only consider whether an implant has failed within 1 year, and disregard the time of death. The BK-CUSUM has a similar problem, where multiple consecutive failures in a short period of time can trigger a false alarm, even if implants fail at reasonable times. A possible example of this can be seen in Figure 1(b). We can see that the many consecutive failures make the chart jump upwards by  $\log(2)$  every time, independent of the probability of failure of those implants at that point in time, thereby rapidly hitting the control limit and afterwards quickly dropping to zero. In contrast to this, the CGR-CUSUM can produce a signal when a few very unlikely failures happen in rapid succession, as can be seen in Figure 1(c). We can see that the Bernoulli CUSUM also almost hits the control limit at a later point, as the upward jumps in the Bernoulli CUSUM chart also depend on the likelihood of failure. Finally, Figure 1(d) shows a hospital which was only detected by the funnel plot. We can see that the hospital experiences a steady stream of failures as the value of the

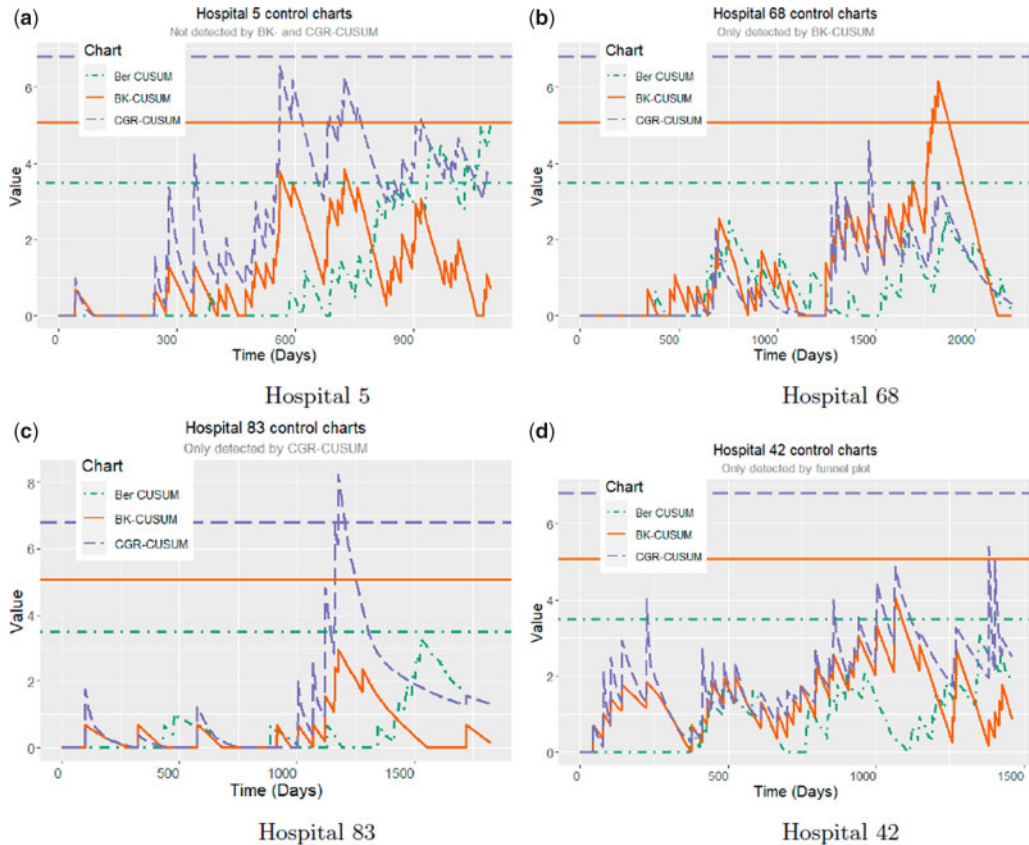


Fig. 1. The (Bernoulli) CUSUM, BK-CUSUM, and CGR-CUSUM charts for four hospitals with their control limits (same color/linetype). The control limits can be found in Table 2. (a) Hospital 5, (b) Hospital 68, (c) Hospital 83, and (d) Hospital 42.

charts is never zero, meaning the proportion of failures at this hospital is reasonably high. The CUSUM charts however indicate that failures are happening at an acceptable rate (possibly slightly higher than target).

The main take-away from this section is that using the continuous time methods it is possible to detect most hospitals signaled by the discrete time methods (much) faster, while guaranteeing a lower percentage of false positive signals. This follows from the fact that the type I error probability for the funnel plot was restricted to approximately 0.05 in 3 years, while we chose control limits for the BK- and CGR-CUSUM such that the probability of a type I error in 6 years was 0.05. Coincidentally, the Bernoulli CUSUM control limit of  $h = 3.5$ , although chosen with a different reasoning, corresponds closely to limiting the type I error in 3 years to 0.05 ( $h = 3.3$ ). Additionally, the results found in this section have to be considered in the correct perspective. We cannot simply state whether the right hospitals were detected at all by any of the charts as we have no information about the true failure rates. For this reason, it is crucial to compare the performance of the charts on a set of hospitals where the true performance of the participating hospitals/implants is known. This will be done in the next section by means of comparing the (average) run length of the procedures under restrictions of the ARL and type I errors when hospitals are performing as expected.

#### 4. SIMULATION STUDIES

In Section 3, we did not know which hospitals were in control. In order to truly compare the performance of the BK- and CGR-CUSUM charts it is crucial to know which hospitals are out of control. This section will compare the methods when the true failure rates at the hospitals are known by means of simulation studies.

In many practical applications, the time to detection after problems occur is of crucial importance in monitoring the quality of the process. Therefore, comparing the performance of inspection schemes in terms of detection speed is important. The expressions found in Section 2 for the approximate average run length of the charts provide a way to compare the charts on a theoretical basis. The equations yielded approximations and depend strongly on the convergence rate of the maximum likelihood estimate. A simulation study can provide a better picture on the finite sample performance of said methods. Besides the detection speed, other quantities such as the type I error and power over time are of interest and will be considered later on in this section.

In Section 3, we chose to impose an upper limit for the MLE  $e^{\hat{\theta}(t)} < 6$  in the CGR-CUSUM, due to the extremely low failure rates in the LROI data. In this section, we also investigate the unrestricted CGR-CUSUM, to investigate the impact of this decision. All simulation studies performed in this section will follow the simulation procedure stated in Section S6 of the [Supplementary material](#) available at *Biostatistics* online.

##### 4.1. A comparison of ARLs

The main goal of this simulation section will be to compare the BK-CUSUM with the new CGR-CUSUM procedure on detection speed for out of control instances. A core assumption of the considered methods is that the change in failure rate happens instantaneously (instead of gradually) and that the true change can be quantified as a fixed increase  $\exp(\theta)$  of the hazard rate. In many practical applications, both assumptions are not likely to hold. We want to examine the effect of wrongly choosing the expected change in failure rate  $\exp(\theta_1)$  in the BK-CUSUM.

To this end, we consider the CGR-CUSUM and two BK-CUSUM procedures with  $\exp(\theta_1) = 1.4$  and 1.8, respectively. We cannot use equal control limits for all charts as this would lead to different properties under the null. For this reason, we determine control limits  $h$  for each procedure such that the in control ( $\exp(\theta) = 1$ ) average run lengths of the procedures are approximately equal to 15 years on a simulated sample size of  $N = 3000$  hospitals. We simulate patient entry by a Poisson process with rate  $\psi = 2.28$  (in days), corresponding to the largest hospitals in the LROI data set. For the in control hazard, we use an exponential distribution with rate  $\lambda = 0.002$  (time in days), so that approximately half of the subjects have failed 1-year postprocedure. The failure rate was chosen to be much higher than in the LROI data set for computational reasons. For this simulation study, risk-adjustment procedures were not considered. For the out of control situation, we want to explore what happens when the chosen  $\theta_1$  is far away from the true value, so we choose true failure rates  $\exp(\theta) \in \{1.2, 1.4, \dots, 3\}$  to generate out of control data sets containing  $N = 3000$  hospitals with the arrival rate and null hazard rate as before.

The run lengths of the two BK-CUSUM procedures are determined for each out of control data set. We determine the run length of the CGI chart on these data sets, giving us an upper bound on the run length of the CGR chart. The results can be found in Table 1, as well as the expected theoretical value of the run length as determined using equations (2.9) and (2.10). The calculation of the Fisher information for the exponential case is discussed in Section S7 of the [Supplementary material](#) available at *Biostatistics* online. A notable result is that at  $\exp(\theta) = 1.4$  the BK-CUSUM with  $\exp(\theta_1) = 1.4$  clearly performs better than the CGR-CUSUM, but the BK-CUSUM with  $\exp(\theta_1) = 1.8$  performs worse than the CGR-CUSUM. This already indicates that the impact of misspecifying  $\theta_1$  can be quite large. Surprisingly, at  $\exp(\theta) = 1.8$

Table 1. Average/median run length, as well as standard deviation and approximate ARL (determined using (2.10) and (2.9)) for two BK-CUSUM with  $\exp(\theta) = 1.4$  and 1.8, as well as the CGR-CUSUM ( $\exp(\theta) = 1$ ) and CGI-CUSUM  $\exp(\theta) > 1$ . Each of the quantities has been determined on a sample of  $N = 3000$  hospitals with hazard ratio  $\exp(\theta)$

$e^\theta$	BK-CUSUM $e^\theta = 1.4, h = 6.82$			BK-CUSUM $e^\theta = 1.8, h = 8.35$			CGR/CGI $h = 7.73$		
	ARL (SD)	MRL	Theory	ARL (SD)	MRL	Theory	ARL (SD)	MRL	Theory
1	5510 (4930)	4056	$\infty$	5478 (4739)	4104	$\infty$	5528 (4666)	4398	$\infty$
1.2	409 (184)	374	1352	639 (366)	572	$\infty$	480 (163)	474	511
1.4	205 (57)	198	227	240 (100)	223	490	229 (72)	228	243
1.6	152 (33)	148	159	153 (48)	145	177	153 (48)	151	162
1.8	127 (24)	125	130	119 (31)	116	128	117 (37)	117	123
2	110 (20)	109	112	101 (23)	99	106	95 (30)	94	100
2.2	99 (16)	98	101	89 (19)	87	92	81 (25)	80	85
2.4	91 (15)	91	92	81 (16)	80	82	71 (23)	71	74
2.6	85 (13)	84	85	74 (14)	73	75	63 (20)	62	65
2.8	79 (12)	79	80	69 (13)	68	70	57 (18)	57	59
3	75 (11)	75	75	65 (12)	64	66	52 (17)	51	54

Table 2. Difference in detection speed (months) of columns with respect to rows. Positive indicating quicker detection and negative indicating slower detection speeds. Values determined on hospitals detected by the funnel plot in the first 3 years, with missing detections omitted

	Median (IQR) difference in detection speed (months) for hospitals detected by funnel plot in the first 3 years			
	Funnel plot $p = 0.95$ yearly	Bernoulli CUSUM $h = 3.5, e^\theta = 2$ monthly	BK-CUSUM $h = 5.1, e^\theta = 2$ monthly	CGR-CUSUM $h = 6.8, e^{\hat{\theta}(t)} \leq 6$ monthly
Funnel plot	0 (0–0)	9 (6–12)	15 (12–17.5)	17 (15–18)
Bernoulli CUSUM		0 (0–0)	7 (4.5–8)	9 (5–10)
BK-CUSUM			0 (0–0)	1 (–0.5 to 3)
CGR-CUSUM				0 (0–0)

the CGR-CUSUM outperforms the other two charts with respect to ARL, but has the largest standard deviation in detection times. In contrast, for small values of  $\theta$  the SD of the BK-CUSUM charts is larger. Finally, for very large values of  $\exp(\theta) > 2$  the CGR-CUSUM seems to be the clear winner. Noticeably, the run lengths of the BK-CUSUM are way more right-skewed than those of the CGR-CUSUM. This can be explained by the nonvariable ( $\theta_1$ ) size of jumps the BK-CUSUM charts can make, in contrast to the variable ( $\hat{\theta}(t)$ ) jump size of the CGR-CUSUM. All in all, we can conclude that with respect to detection speed the BK-CUSUM is the preferred chart when the true hazard ratio is small ( $\exp(\theta) \leq 1.4$ ) and/or we have a lot of confidence in our prior knowledge. The approximate average run lengths determined using (2.10) and (2.9) seem to work quite well both for the BK-CUSUM as well as for the CGR-CUSUM, especially for large ( $\exp(\theta) > 1.2$ ) true hazard ratios.

These simulation results give rise to the presumption that the CGR-CUSUM should perform better when the rate of failure is variable, especially combined with large values of  $\theta$ . This is also what we saw in Section 3, when we applied the CGR-CUSUM to a real-life data set.



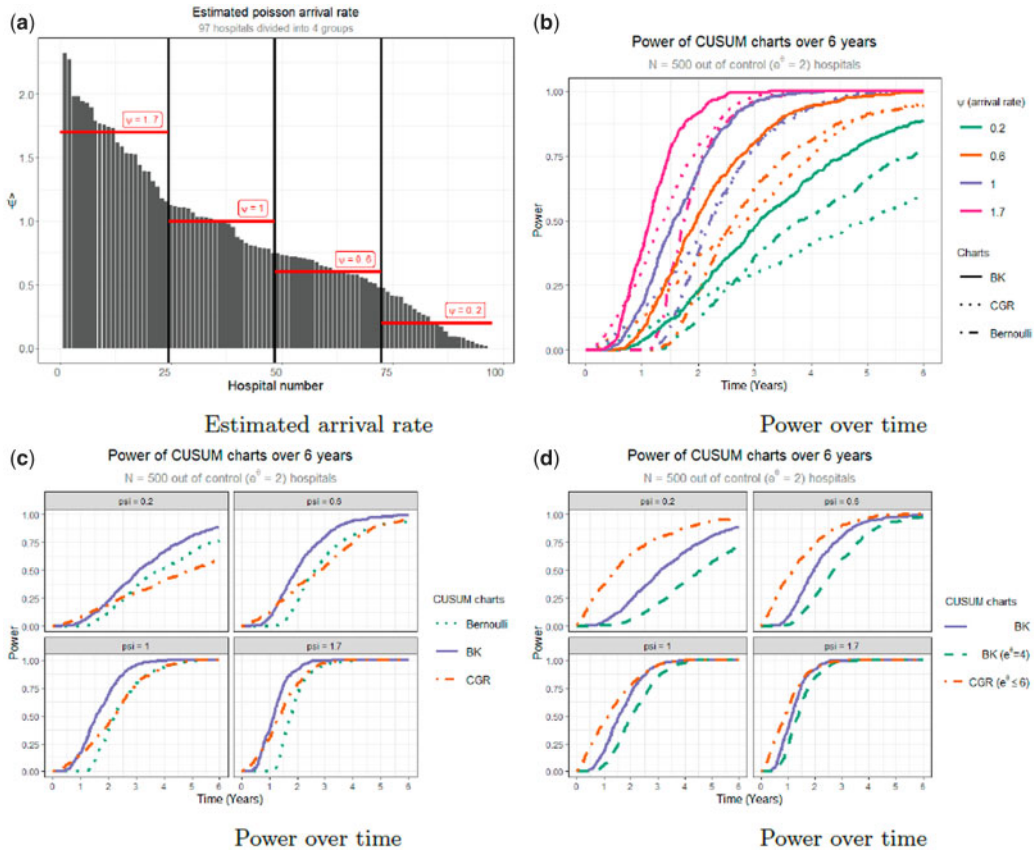


Fig. 2. (a) Estimated arrival rate as well as the subdivision of the hospitals into four groups. (b) Simulated power of the Bernoulli and continuous time CUSUM charts on a sample size of  $N = 500$  out of control ( $e^{\theta_1} = 2$ ) hospitals using control limit values such that the simulated in control type I error  $\alpha \approx 0.1$  in 6 years (see Table 3). (c) Figure (b) faceted over the different values of  $\psi$ . (d) Comparison of the power over time of two BK-CUSUM charts ( $e^{\theta_1} \in \{2, 4\}$ ) and the CGR-CUSUM with  $e^{\theta_1} \leq 6$ .

#### 4.2. Power under type I error restriction

Instead of restricting the in control ARL, Biswas and Kalbfleisch (2008) and Begun and others (2019) have chosen to restrict the simulated in control type I error to 0.15 in 5 years and 0.1 in 8 years respectively. Besides this, hospitals vary in size and therefore the number of patients treated per day. This difference in patients treated per time unit, in our model expressed by the parameter  $\psi$ , has a strong influence on the detection speed and power of the procedures.

For this reason, in this section, we will determine the power over time of two BK-CUSUM procedures ( $e^{\theta_1} \in \{2, 4\}$ ), two CGR-CUSUM ( $e^{\theta_1} \leq \{6, 8\}$ ) procedures and the Bernoulli CUSUM ( $e^{\theta_1} = 2$ ) for hospitals of different sizes under a restricted type I error. We consider four groups of hospitals by size, with  $\psi \in \{0.2, 0.6, 1, 1.7\}$ . These values were determined by subdividing the hospitals in the LROI data set into four groups by size and averaging over their estimated patient arrival rate, see Figure 2(a). Using the simulation procedure in Section S6 of the Supplementary material available at Biostatistics online with resampling, we find control limits for all considered methods by limiting the risk-adjusted simulated type I error in 6 years to  $\alpha \approx 0.1$  on  $N = 500$  in control hospitals, see Table 3. Note that the control limits

Table 3. Control limits determined on a sample size of  $N = 500$  in control ( $e^\theta = 1$ ) hospitals such that the type I error in 6 years  $\alpha \approx 0.1$

$\psi$	Control limit h				
	Bernoulli CUSUM $e^{\theta_1} = 2$	BK-CUSUM $e^{\theta_1} = 2$	BK-CUSUM $e^{\theta_1} = 4$	CGR-CUSUM N/A	CGR-CUSUM $e^{\hat{\theta}(t)} \leq 6$
0.2	2.62	3.15	4.64	7.31	4.68
0.6	3.71	4.19	5.81	7.73	5.79
1	4.34	4.76	6.34	8.27	6.51
1.7	4.72	5.41	6.79	8.54	6.69

for the unrestricted CGR-CUSUM are very close together for all values of  $\psi$ . This is a consequence of no longer bounding the MLE  $\hat{\theta}(t)$  from above.

We then simulate  $N = 500$  out of control ( $\exp(\theta) = 2$ ) hospitals for each considered value of  $\psi$ . The detection times on these data sets are then determined for each chart using the control limits in Table 3. The resulting power over time for the BK-CUSUM ( $e^{\theta_1} = 2$ ), the Bernoulli CUSUM and CGR-CUSUM can be seen in Figures 2(b) and (c). The BK-CUSUM with correctly specified parameters clearly has the best power over time for hospitals of all sizes. The CGR-CUSUM performs worse than the Bernoulli CUSUM for low arrival rates, but does better as the arrival rate increases. This is due to the very high value of the control limit for the CGR-CUSUM, causing detections to be delayed.

We also compare the power over time of the BK-CUSUM ( $e^\theta = 2$ ) with that of the CGR-CUSUM ( $e^{\hat{\theta}(t)} \leq 6$ ) and the BK-CUSUM ( $e^\theta = 4$ ) in Figure 2(d). In this figure, the CGR-CUSUM is clearly the winner for all hospital sizes. The control limits for the restricted CGR-CUSUM are much smaller than for the unrestricted CGR-CUSUM. This is because the unrestricted CGR-CUSUM can produce extremely large estimates for  $\hat{\theta}(t)$ , therefore becoming very unstable even in the in control situation. The BK-CUSUM ( $e^{\theta_1} = 4$ ) with incorrectly specified parameters performs the worst for all hospital sizes. Notably, all three procedures seem to converge towards the same power over time graph as the arrival rate increases, which was not the case in Figure 2(c). We conclude that the CGR-CUSUM can yield the best power over time, but depending on the nature of the data restricting the value of  $\hat{\theta}(t)$  might be necessary to achieve such a performance.

## 5. DISCUSSION

For almost all applications, the CGR-CUSUM will yield earlier detection times than the BK-CUSUM since in general the change of failure rate at a hospital will not be of a fixed size and will happen gradually instead of instantaneously. For this reason, the CGR-CUSUM will perform better in practical scenarios where the expected true hazard ratio  $\exp(\theta)$  is not known in advance or variable over time. This was demonstrated in our application of the charts to the LROI data set. The application of the BK- and CGR-CUSUM charts on the LROI data set also showed that in practice the CGR-CUSUM outperforms the BK-CUSUM with respect to detection times, while retaining a similar number of “false” detections. It is important to note that we do not know whether the hospitals detected by the funnel plot were the hospitals with “true” problems, instead operating in line with [van Schie and others \(2020\)](#) by taking the funnel plot as the golden standard. We cannot be sure that the chosen expected hazard ratio for the BK-CUSUM was in line with reality. All in all, we can conclude that the CGR-CUSUM is the preferred method for quality inspection, especially for large arrival rate  $\psi$ . From the simulation study in Section 4.2, we concluded that the CGR-CUSUM can yield better power than the BK-CUSUM, but might require appropriately



restricting the values of  $\hat{\theta}(t)$ . Even though the CGR-CUSUM was created with the goal of specifying fewer parameters, we believe that bounding  $\hat{\theta}(t)$  is often more tractable than correctly specifying the expected hazard ratio. This restriction was necessary for the LROI data, but not all survival data will have an extremely low failure rate and therefore the CGR-CUSUM could also perform well without this restriction, as was seen in Section 4.1.

### 5.1. Recommendations for practice

For practical applications, we suggest using the CGR-CUSUM for quality control, keeping in mind that it restricting the maximum likelihood estimate to an appropriate range might be necessary (i.e.,  $e^{\hat{\theta}(t)} \leq 6$ ). For small volume hospitals the BK-CUSUM could be preferred, as long as there is some prior information about the expected increase in failure rate. This way the small amount of information retained from patients can be partly compensated by prior knowledge. The use of a funnel plot is not advised as it is not a real-time procedure and has the potential disadvantage of an increased risk of a type I error incurred by performing a multiple testing procedure.

We advise determining control limits  $h$  for CUSUM charts either by restricting the simulated probability of a type I error over a time frame or by restricting the in control average run length of the charts. The first method may be preferred due to the lower computational requirements.

Ideally, the baseline hazard rate should be determined on a data set which is known to be in control. Realistically, this is unlikely to be feasible in many applications. The practice of considering the national average rate of failure to be in control is often sufficient. An important consideration is that any major change in the distribution of risk factors in the population will require a recalculation of control limits. Whereas information on the failure of patients can be collected in real time, the aggregation of such data over multiple hospitals is not likely to happen in real time. If the risk distribution has changed over this frame of time, it might be necessary to reconstruct the CUSUM charts, possibly leading to new or different detections.

### 5.2. Limitations

In the considered model, we assume that observations can only be right censored. This is because in the setting of arthroplasty surgery left and interval censoring are of little interest. The same is not true for competing risks mechanisms. [Begun and others \(2019\)](#) have considered a similar procedure to [Biswas and Kalbfleisch \(2008\)](#) with the addition of frailty terms and competing risks, allowing for dependent competing risks. Even though they could not find an indication that the competing risks of death and revision surgery are dependent in their data, their methods can be carried over to our procedure as well. Should we be interested in detecting a decrease in the rate of failure using the BK- or CGR-CUSUM, a two-sided procedure as suggested by [Page \(1954\)](#) can be considered where the hypotheses of  $\theta = 0$  against  $\theta = \theta_1 < 0$  are used for constructing the likelihood ratio. This yields the CUSUM charts with switched positive and negative signs.

### 5.3. Future work

Additions to the CGR-CUSUM and BK-CUSUM should be considered. Notably, the power of the unrestricted CGR-CUSUM was lacking for hospitals with a low volume of patients. This is largely due to the (relatively) very high value of the control limit of the CGR-CUSUM (see Table 3). These values are so high because the CGR-CUSUM will often have an initial spike upwards when the first failure is observed due to the maximization over previous patients (i.e., all patients before the first failed patient are ignored). When the volume of patients or failure rate is low, this leads to a large uncertainty in the determination of the MLE  $\hat{\theta}(t)$ . To counteract this, in Section 3, we introduced the upper limit  $\hat{\theta}(t) \leq \ln(6)$ .

Another solution would be to impose a time-dependent control limit which is large at the start of the study and decreases until it reaches a fixed value, allowing the CGR-CUSUM to converge before yielding detections. A patient shuffling or weighing mechanism can be added to the CGR-CUSUM chart in order to yield quicker detection in the case of clustered failures in the past. For this, the mechanisms used by Steiner and Jones (2009) and Grigg (2018) can be used as inspiration. Finally, a mechanism where patients have periods when they are not at risk of failure can be incorporated into the chart as well.

## 6. SOFTWARE

Software in the form of an R package, together with a sample input data set and complete documentation are available on CRAN at <https://cran.r-project.org/package=success>. Code to reproduce the results in this article can be found on GitHub at [https://github.com/d-gomon/success\\_example](https://github.com/d-gomon/success_example).

## SUPPLEMENTARY MATERIAL

Supplementary material is available online at <http://biostatistics.oxfordjournals.org>.

## ACKNOWLEDGMENTS

The Dutch Arthroplasty Register (LROI) is gratefully acknowledged for providing their arthroplasty database, under agreement LROI 2020-053. We wish to thank the associate editor and referees for their help in improving this manuscript.

*Conflict of Interest:* None declared.

## FUNDING

Dutch Research Council (NWO) Veni 192.087.

## REFERENCES

- AALLEN, O. O., BORGAN, Ø. AND GJESSING, S. (2008). *Survival and Event History Analysis: A Process Point of View*, 1st edition. New York, NY: Springer.
- BEGUN, A., KULINSKAYA, E. AND MACGREGOR, A. J. (2019). Risk-adjusted CUSUM control charts for shared frailty survival models with application to hip replacement outcomes: a study using the NJR dataset. *BMC Medical Research Methodology* **19**, 217.
- BISWAS, P. AND KALBFLEISCH, J. D. (2008). A risk-adjusted CUSUM in continuous time based on the Cox model. *Statistics in Medicine* **27**, 3382–3406.
- DUTCH ARTHROPLASTY REGISTER (LROI). (2020). Online LROI annual report 2020. <https://www.lroi-report.nl/previous-reports/online-lroi-report-2020/>.
- GRIGG, O. A. (2018). The STRAND chart: a survival time control chart. *Statistics in Medicine* **38**, 1651–1661.
- HAN, D. AND TSUNG, F. (2007). Detection and diagnosis of unknown abrupt changes using CUSUM Multi-Chart Schemes. *Sequential Analysis* **26**, 225–249.
- PAGE, E. S. (1954). Continuous inspection schemes. *Biometrika* **41**, 100–115.
- SEGO, L. H., REYNOLDS, M. R. AND WOODALL, W. H. (2009). Risk-adjusted monitoring of survival times. *Statistics in Medicine* **28**, 1386–1401.
- SIEGMUND, D. AND VENKATRAMAN, E. S. (1995). Using the generalized likelihood ratio statistic for sequential detection of a change-point. *The Annals of Statistics* **23**, 255–271.

- SPIEGELHALTER, D. J. (2005). Funnel plots for comparing institutional performance. *Statistics in Medicine* **24**, 1185–1202.
- STEINER, S. H., COOK, R. J., FAREWELL, V. T. AND TREASURE, T. (2000). Monitoring surgical performance using risk-adjusted cumulative sum charts. *Biostatistics* **1**, 441–452.
- STEINER, S. H. AND JONES, M. (2009). Risk-adjusted survival time monitoring with an updating exponentially weighted moving average (EWMA) control chart. *Statistics in Medicine* **29**, 444–454.
- THERNEAU, T. M. (2020). *A Package for Survival Analysis in R*. R package version 3.2-7 <https://CRAN.R-project.org/package=survival>.
- VAN BUUREN, S. AND GROOTHUIS-ODUSHOORN, K. (2011). mice: multivariate imputation by chained equations in R. *Journal of Statistical Software* **45**, 1–67.
- VAN SCHIE, P., VAN BODEGOM-VOS, L., VAN STEENBERGEN, L. N., NELISSEN, R. G. H. H. AND VAN DE MHEEN, P. J. MARANG. (2020). Monitoring hospital performance with statistical process control after total hip and knee arthroplasty. *Journal of Bone and Joint Surgery* **102**, 2087–2094.
- VAN STEENBERGEN, L. N., DENISSEN, G. A. W., SPOOREN, A., VAN ROODEN, S. M., VAN OOSTERHOUT, F. J., MORRENHOF, J. W. AND NELISSEN, R. G. H. H. (2015). More than 95% completeness of reported procedures in the population-based Dutch Arthroplasty Register: external validation of 311,890 procedures. *Acta Orthopaedica* **86**, 498–505.

[Received December 7, 2021; revised August 23, 2022; accepted for publication August 29, 2022]