Continuous time control charts: Generalizations, simulation studies and an application to the Dutch Arthroplasty Register (LROI)

Gomon, D.

Continuous time control charts: generalizations, simulation studies and an application to the Dutch Arthroplasty Register (LROI)

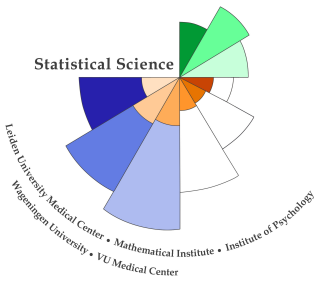
D. Gomon

Thesis supervisor: dr. S.L. van der Pas

Master thesis

April, 2021

Specialization: Data Science



MATHEMATICS
&
STATISTICAL SCIENCE
FOR THE LIFE AND BEHAVIOURAL SCIENCES

Leiden University
Mathematical Institute

Foreword

This thesis was written in support of graduation for the MSc. Mathematics and MSc. Statistical Science programs at Leiden University.

This project was undertaken in order to improve the quality of transplantation surgery performed in the Netherlands. The first task was to write a research application form in order to get approval from the LROI [1] to use their data for this purpose. Luckily, our application was approved. To write this proposal we started with researching relevant literature in the field of medical statistics. We decided to base our research on an existing method for medical monitoring: a CUSUM monitoring scheme developed by Biswas & Kalbfleisch [2]. The main theoretical contribution towards the field is the development of two generalizations of this method, which we call the continuous time (maximized) generalized likelihood ratio charts. A simulation study was devised to assess these new charts and compare them with the existing method. To ensure feasible computation times, quite a lot of code optimisation had to be performed.

Afterwards, a more practical direction was taken. The available data set was analysed and cleaned. A decision was made about which performance measures are of interest. The previously mentioned methods were applied to the data. To compare with current practices at the LROI, their current inspection scheme was also considered. Another simulation study was thought out and performed in order to simulate a realistic scenario.

Finally, the core findings were reported in this thesis. In the near future, we will discuss our findings with the LROI in order to give recommendations for further monitoring of transplantation surgery.

The R code used for this thesis can be requested from Daniel Gomon by sending an e-mail to daniel.gomon@hotmail.com.

Acknowledgements

First and foremost I would like to thank my thesis supervisor Stéphanie who has looked out for me before I even started this project. Having only once mentioned my selfish desire to combine two theses into one, Stéphanie approached me a few weeks afterwards with a suitable topic and a well laid out plan allowing for just this. Besides this, I have received excellent guidance. There was not a moment where it was unclear to me what I should do, and whenever I was stuck I would receive help within a day's time.
Big thanks to my friends Onno and Geerten for giving me general guidelines on my thesis.
I am thankful for the work performed by Caroline Kok on the subject of CUSUM charts and their generalizations. Her work allowed me to quickly attain all necessary knowledge on this method, sparing me at least a few weeks of work.
My parents were the ones who introduced me to mathematics and deserve a big thank you, not only for making this possible, but also for showing me a direction in life which allows me to support myself while doing something I truly enjoy.
I would like to thank Bart for distracting me from work when I was at my most productive time of day.
Finally, I would like to especially thank Barthold, Wing-Tsi and Fabienne for reading and providing feedback on the core parts of my thesis, even though they are not familiar with statistics. Besides this, I would like to thank Fabienne for her emotional support throughout this process.

Contents

Abstract

In recent years, monitoring the quality of medical care has become of interest in multiple countries. An example is the Dutch Arthroplasty Register (LROI) [1], which keeps records of all joint transplantation surgery performed in the Netherlands. In order to contribute to the monitoring of transplantation surgery, the main goal of this thesis is to detect an increase in the failure rate of prostheses at a hospital as soon as possible.

In light of this mission, we consider a current procedure employed for this purpose: the Bernoulli cumulative sum (CUSUM) chart [3]. After exploring several properties of this method, we argue that it is not optimal for the problem at hand due to the survival nature of the data. For this reason we consider a risk-adjusted continuous time extension of the CUSUM for survival outcomes (CTCUSUM), developed by Biswas & Kalbfleisch [2].

The CTCUSUM chart requires an assumption as to the degree of a future increase in failure rate. The assumption is made by choosing a parameter in advance, requiring prior knowledge about the problem at hand. To mitigate this need of prior knowledge (which is often not available), we propose two generalizations of the CTCUSUM named the continuous time (maximized) generalized likelihood ratio (CT(MAX)GLR) charts. A theoretical approximation is made to the average run length (average time to detection) for both these charts. Conveniently, this approximation also yields an expression for the average run length of the CTCUSUM chart. Finally, theoretical bounds are found for the asymptotic expected value of the CT(MAX)GLR chart. The obtained theoretical quantities are assessed by means of a simulation study. The loss in detection speed for not, or incorrectly, choosing the expected increase in failure rate is discussed.

The previously mentioned charts are applied to a data set from the LROI and compared with the current procedure in place at the LROI: funnel plots. We find that detection times can be greatly improved by employing the described continuous time charts, yielding up to 15 months (median) earlier signalling of problematic hospitals. Since no information is available on whether hospitals experienced a drop in their quality of care, an additional simulation study is performed where the power, sensitivity and specificity of the CTCUSUM and Bernoulli CUSUM is compared. The CTMAXGLR is not included in the simulation study, as it is very computationally demanding. Finally, recommendations are made for further monitoring of arthroplasty outcomes specifically and for practitioners in general.

Notation

Patient Characteristics	
X_i	Survival time of patient i
S_i	Chronological time of entry into study of patient i
T_i	Chronological time of failure of patient i
C_i	Chronological time of right-censoring of patient i
Z_i	Covariate vector of patient i
Symbols	
h	Control limit for control charts
θ_0, θ_1	Null and alternative hypothesis values for CUSUM charts ($\theta = \theta_1$)
θ	Either the value of θ chosen for a CUSUM chart, or θ_1
μ	True value of the relevant parameter
ψ	Poisson arrival rate (daily or yearly)
β	Risk-adjustment coefficient vector
C	Constant indicating up until how many years post primary procedure a failure is qualifying
t	Value of (relevant) time
n	Number of observations (total or at relevant time t)
$\hat{\theta}_t$	Maximum likelihood estimate of θ at time t, see 6.2.1
γ_u	Defined in (39)
Functions	
$N_i^D(t)$	Failure indicator for patient i at time t
$N^D(t)$	Amount of failed patients at some hospital D at time t ($=\sum_i N_i^D(t)$)
$h_0(t)$	Non risk-adjusted baseline hazard rate at time t
$h_i^\theta(t)$	Risk-adjusted hazard rate at time t for patient i with covariates Z_i and multiplied by e^θ ($= h_0(t)e^{\beta Z_i}e^\theta$)
$H_0(t)$	Non risk-adjusted cumulative baseline hazard at time t
$H_i^\theta(t)$	Risk-adjusted cumulative baseline hazard at time t for patient i with covariates Z_i and multiplied by e^θ ($= H_0(t)e^{\beta Z_i}e^\theta$)
$Y_i(t)$	Indicator whether patient i is active at time t
$\Lambda_i^\theta(t)$	Cumulative intensity of patient i at time t with excess risk factor e^θ
$A(t)$	Summed baseline cumulative intensity at some hospital ($= \sum_i \Lambda_i(t)$)
$I(\theta, t)$	Fisher information in all observations at time t
$\overline{I(\theta, t)}$	Fisher information in all observations at time t, divided by amount of total observations n at time t
Acronyms	
ARL	Average run length
SPRT	Sequential probability ratio test
CUSUM	Cumulative sum (chart)
GLR	Generalized likelihood ratio (chart)
MAXGLR	Maximized generalized likelihood ratio (chart)
CT(XXX)	Continuous time (XXX) (chart)
ML(E)	Maximum likelihood (estimate)
RA	Risk-adjustment
Cox PH	Cox proportional hazards
Terms	
(in)active	A patient is active after entering into the study up until time of failure or censoring, otherwise a patient is inactive.
discrete time CUSUM	Cumulative sum chart as in section 4.2
discrete time GLR	Generalized likelihood ratio chart as in section 4.3
Bernoulli CUSUM	Cumulative sum chart for Bernoulli outcomes

1 Introduction

A cumulative sum chart (CUSUM) is a chart commonly used in medical fields to determine whether the quality of treatment has diminished (or improved) over time. Despite this, the CUSUM does not originate from the medical field, but from the 20th century industrial revolution. During this industrial revolution many processes involving machinery had to be monitored ensuring that the production was operating efficiently, thereby reducing waste [4]. Consequently, the CUSUM chart was not developed with survival outcomes in mind. For medical purposes however, survival outcomes are of great interest, as often the duration of a patient's survival after their treatment begins is observed. Due to this, current medical CUSUM inspection schemes are forced to consider whether patients have died C years after their primary procedure. By doing so, the information provided by the survival or death of said patient are ignored until C years post procedure. To overcome this problem, continuous time extensions of the CUSUM will be the core topic of this thesis. In these continuous time charts patients are considered at all times, incorporating the information provided by the survival of patients at any time. The goal is to improve on the detection speed and power of the CUSUM chart in monitoring the quality of transplantation surgery at the Dutch Arthroplasty Register (LROI [1]) by considering these continuous time charts, thereby increasing the quality of transplantation surgery in the Netherlands.

1.1 The cumulative sum chart

An important step towards the discovery of CUSUM charts was the development of the sequential probability ratio test (SPRT) in 1945 by Wald [5]. The SPRT can be used to test the hypotheses $(H_0 : \mu = \theta_0, H_1 : \mu = \theta_1)$ whether the rate (parameter) of a process has changed, for example by testing whether a cereal production line puts the right amount of cereal into the box [4]. Using the SPRT one can conclude that the process is in control (operating efficiently) or out of control (operating inefficiently), accepting or rejecting the null hypothesis respectively.

Later, Page [3] extended the SPRT to another setting. In this setting the process was assumed to be in control until a signal was produced indicating that the process was out of control. Under these assumptions, this method could be used to continuously monitor for an unfavourable change in parameter. Page named this statistical procedure the CUSUM.

A CUSUM chart for detecting an increase in parameter is shown in Figure 1. The procedure is quite simple: any unfavourable outcome (for example death) leads to a rise in the value of the chart, depending on the likelihood of observing this outcome for said individual. This likelihood might depend on some of their properties (called covariates). Any favourable outcome (such as remission) leads to a decrease in the value of the chart, depending on some covariates. In Figure 1 the chart reaches zero at observation four. With a SPRT we would conclude that the process is in control. On the other hand, a CUSUM chart is aimed at detecting a decrease in quality. Therefore, we keep observing outcomes until the control limit (denoted by h) is passed. In Figure 1, the CUSUM chart crosses the control limit at observation seven, signalling a decrease in quality. It is also possible to simultaneously construct a CUSUM chart for detecting an increase in quality, which can be plotted besides (mirrored underneath) the current chart. This will not be of interest in this thesis.

Finally, in the late 1990's the case of Harold Shipman, a general practitioner who is thought to have murdered up to 250 of his patients without being detected, was the cause of outrage in a town in the United Kingdom. Spiegelhalter et al. [6] argued that Shipman's malpractice could have been detected substantially sooner, had a CUSUM monitoring scheme been in place. Besides criminal medicine, quality control in many other medicinal applications is rapidly gaining in popularity. As an example, the Scientific Registry of Transplant Recipients [7] in the US uses monthly CUSUM charts

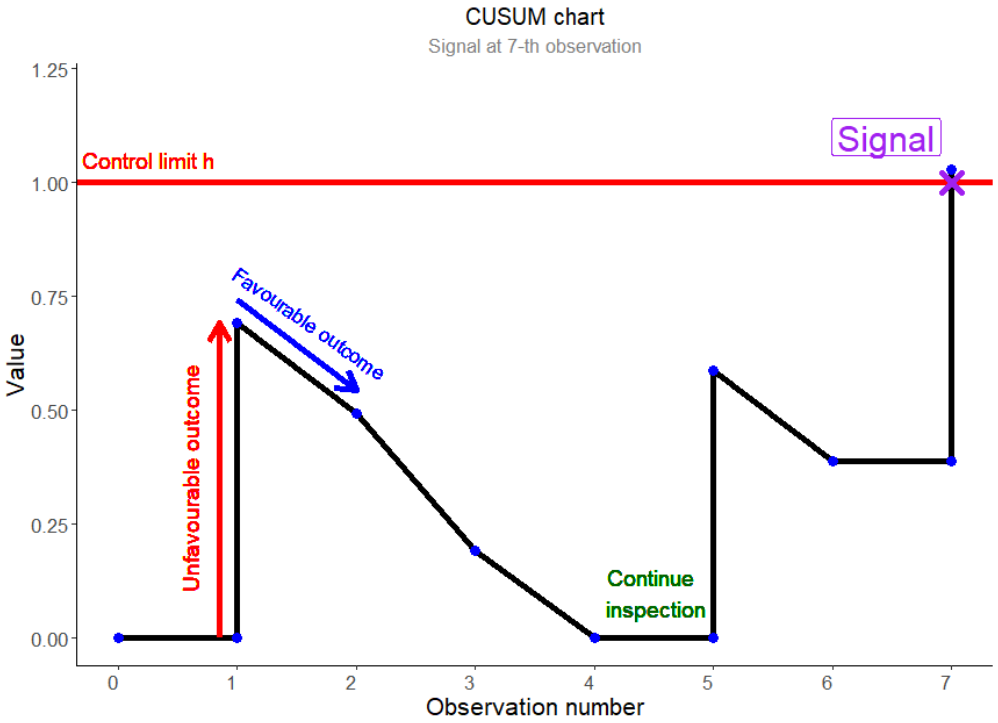


Figure 1: Example of a CUSUM chart. An unfavourable outcome (i.e. death, complications) increases the value of the chart, while a favourable outcome (i.e. remission) decreases the value of the chart. The chart produces a signal when it crosses the value of the control limit h, which is chosen in advance.

in its assessment of almost all transplant surgeries.

1.2 Drawbacks of the CUSUM in medical monitoring

As the CUSUM is used to test a point hypothesis of a change in parameter, one of the biggest drawbacks in medical monitoring is the necessity to quantify the expected loss in the quality of care by choice of parameter values. Aside from choosing some required standard of care (in the form of θ_0), it is often unclear at what threshold the quality of care is no longer acceptable, complicating the choice of θ_1. In addition, the choice of control limit h directly influences the sensitivity, specificity and detection speed of the chart.

Besides the mentioned complication, many applications in medicine involve survival outcomes. With survival outcomes, the survival of the patient after the first treatment (procedure) is of interest. The CUSUM chart described above can only be constructed at times when an outcome is observed. Instead of waiting for the failure time of each patient, current CUSUM charts in medical monitoring consider whether patients have failed C years post primary treatment. This procedure is called the Bernoulli CUSUM. The necessity to dichotomize the outcomes after C years comes from the low failure rate of patients in most medical applications, as will be the case in our data set of interest (1.7% in the first year post procedure). For arthroplasty outcomes, $C = 1$ is chosen most of the time [2] [8], meaning that a patient is only considered 1 year after the procedure has passed. Consequently, the information provided by their survival or death up until that point is ignored. On top of this, any failure past the one year mark will not be incorporated into the value of the chart, possibly leading to detection delays.

1.3 Main goal of thesis

The main goal of this thesis is to increase the practical utility of CUSUM charts, by tackling two of their inherent problems:

1. The need to quantify the expected increase in failure rate (choice of θ_1 in advance).
2. The inability to continuously incorporate survival information into the chart.

In practice this means that in the improved methods every patient will be able to continuously contribute to the value of the chart as well as removing the need to specify an expected increase in failure rate, thereby allowing for the detection of any kind of deterioration. These improvements will be achieved by combining the work of Wald [5] and Biswas & Kalbfleisch [2] into two chart, which wil be called the continuous time (maximized) generalized likelihood ratio (CT(MAX)GLR) charts. The CTGLR can be used to test for an immediate change in failure rate, while the CTMAXGLR can be used to test for a delayed change in failure rate (similar to the CUSUM procedure).

Wald has developed a chart which Kok [9] calls the maximized generalized likelihood ratio test, which "automatically" selects a value for the alternative by means of a maximum likelihood estimate. Biswas & Kalbfleisch [2] combined their knowledge of survival analysis with the theory of CUSUM charts to arrive at a continuous time CUSUM chart (CTCUSUM). The CTCUSUM implements the knowledge that an individual has not failed yet into its value at any given time, aimed at increasing the power and detection speed of the chart. An example of such a chart is given in Figure 2. Contrary to the discrete time CUSUM chart, the CTCUSUM chart always rises by the chosen value of θ_1 when a failure is observed. Afterwards, the chart drifts downwards depending on the amount of people at risk, and their likelihood of failure in that timeframe (depending on some covariates). Combining these two properties, the CT(MAX)GLR chart automatically selects an appropriate value for θ_1, as well as incorporating survival outcomes into the chart.

1.4 Results and recommendations

The main theoretical result of this thesis is the asymptotic distribution of the CTGLR chart in section 6, which allows us to find an upper bound for the out of control asymptotic average run length of the CTMAXGLR and CTCUSUM charts, indicating how quickly a signal can be expected under the alternative hypothesis (of an increase in failure rate). In section 6.10, we show that asymptotically the CTMAXGLR enjoys quicker detection times when the expected increase in failure rate (θ_1) is chosen wrongly in the CTCUSUM chart. By means of simulation we confirm this finding. Asymptotic bounds for the expected value of the CT(MAX)GLR chart are found and assessed using a simulation study in section 7.

Having found these results, a more practical direction is taken in section 12. We apply the CTCUSUM and CTMAXGLR to a data set from the Dutch Arthoplasty Register (LROI) [1]. This data set contains information about hip replacement surgeries in hospitals all over the Netherlands. We compare these charts with respect to the Bernoulli CUSUM and funnel plot currently employed by the LROI [10] [11]. We find that major improvements in detection speed can be achieved by employing the new charts. In the first three years of data the Bernoulli CUSUM leads to a median 9 months faster detection compared to the funnel plot, improved by the CTCUSUM by another 6 months (median) and on top of that improved by the CTMAXGLR by 5 months (median). Unfortunately, these come at the cost of possible false detections past the three year margin. As we lack information on the true rate of failure at hospitals, it remains unclear how many false detections any of the charts produce. By means of simulation we compare the power of the CTCUSUM with that of the Bernoulli CUSUM in section 13, as well as comparing them on sensitivity under restrictions on their specificity.

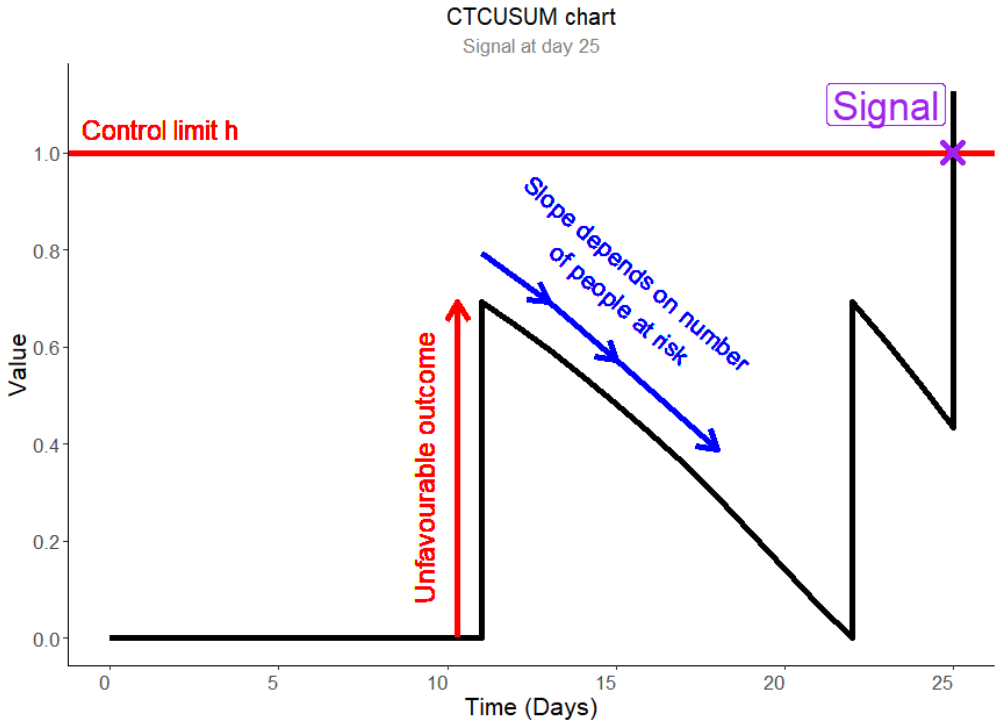


Figure 2: Example of a CTCUSUM chart. An unfavourable outcome (i.e. death, failure) increases the value of the chart. The chart drifts downwards depending on the amount of people at risk, and their likelihood of failure at that time. The chart produces a signal when it crosses the value of the control limit h, which is chosen in advance.

The CTMAXGLR is left out of consideration due to its demanding computational requirements. We conclude that the CTCUSUM is the more powerful chart with quicker detection times, especially for hospitals with a small number of patients per time unit.

Finally, to assess the false detection rate of the (continuous time) CUSUM charts, we perform a "realistic" simulation study in section 13.4 where the true rate of failure at hospitals is known. Drawing from the result of this study, recommendations are made towards further monitoring of arthroplasty outcomes in section 14.1.

1.5 Reading guide

This thesis is split into two parts:

1. Part I focusses on the theoretical development of the CTCUSUM and CTMAXGLR charts, deriving the charts and proving some of their properties. Sections 2 and 3 introduce some basic theory necessary for later proofs and understanding. Section 4 introduces the reader to the theory available on likelihood ratio tests for discrete time outcomes (such as the SPRT and CUSUM). Section 5 explores the CTCUSUM developed by Biswas & Kalbfleisch [2]. The CTGLR and CTMAXGLR are derived in section 6, as well as some of their properties. Finally, in section 7 a simulation study is performed to assess the obtained results and a conclusion is formulated in section 8.

2. Part II introduces the funnel plot currently employed by the LROI in section 9. Then the theory from part I about CUSUM and CTCUSUM/CTMAXGLR charts is summarized in sections 10 and 11 respectively. These methods are applied to the data set and compared on various indicators in section 12. Afterwards, a simulation study on the basis of this data set is performed in section 13. Final conclusions and recommendations for practice are then given in section 14.

Readers from the **Mathematics committee** should read the whole of part I, meaning sections 2 - 8. If the reader is interested in a practical application, part II is also recommended.
Readers from the **Statistical Science committee** should read section 2 and skip the rest of part I to continue in part II, meaning sections 9 - 14. If the reader is interested in the specifics of the shown properties for the charts, the rest of part I is recommended.

Part I
Mathematical theory and simulations

In this part of the thesis we first introduce some theory on survival analysis in section 2, followed by a small amount of theory about renewal processes and queueing theory in section 3, before arriving at section 4, where we present and shed some light on a few likelihood ratio tests. These three sections introduce the reader into the theory necessary to understand the CTCUSUM developed by Biswas & Kalbfleisch [2], which is introduced in section 5. Their method can be used to sequentially test for a change in the rate of some statistical process, and will form the basis for our further developments. These first few chapters can be seen as a reiteration of relevant existing theory in the field.

In section 6 we present, to the best of our knowledge, a new method called the CT(MAX)GLR which is based on the CTCUSUM. Besides this, we show some (asymptotic) properties of the CT(MAX)GLR. Finally, we compare the theoretical results of our method with the theoretical results obtained by Biswas & Kalbfleisch [2]. To asses whether the developed theory can be applied in practice, some simulations are performed in section 7. Here we primarily assess how the CT(MAX)GLR compares to the CTCUSUM in terms of detection speed, under the assumptions of the models. Besides this we determine whether the assumptions made to develop the theory are realistic and confirm our asymptotic results.

Whereas part I of this thesis focusses on our theoretical developments in the field, part II will focus on applying the developed methods to a real-life data set and comparing the results with that of methods currently employed to monitor arthroplasty outcomes and other medical data sets involving survival outcomes.

2 Survival analysis

Some background theory on survival analysis will be introduced in this section. This theory will be based on the book written by Klein & Moeschberber [12]. It will be key in the transformation of the discrete time CUSUM (section 4.2) into the continuous time CUSUM (section 5) and finally into the continuous time (MAX)GLR (section 6), which is one of the main results of this thesis.

2.1 Survival function

In survival analysis we consider the time until some event of interest occurs, which as the name 'survival' implies is often the death of an individual. In statistics we often assume a certain underlying outcome distribution for the observed variables. Suppose the time to death X has a certain underlying distribution F and density function f. In this thesis we will restrict ourselves to continuous survival times. Sometimes we are more interested in the **survival function** S given by:

$$S(t) = \mathbb{P}(X > t) = 1 - F(t) = \int_t^\infty f(s)ds.$$

This quantity tells us the probability that an individual will survive beyond time t, which is a continuous strictly decreasing function. From the equality above we recover the relationship:

$$f(t) = -\frac{dS(t)}{dt}.$$

2.2 Hazard function

An important quantity throughout this thesis is the **hazard function** h, also known as the **hazard rate**, which is not to be confused with the hazard ratio. It is given by:

$$h(t) = \frac{f(t)}{S(t)} = -\frac{d\ln(S(t))}{dt} = \lim_{\Delta t \to 0} \frac{\mathbb{P}(t \leq X \leq t + \Delta t | X > t)}{\Delta t}. \tag{1}$$

The hazard function is therefore the probability of instantaneous death given that the individual has survived to see present time, divided by the limit of an infinitesimally small time difference to zero. The hazard function must be non-negative. Additionally we define the **cumulative hazard function** H as:

$$H(t) = \int_0^t h(s)ds = -\ln(S(t)).$$

This allows us to recover the survival function as follows:

$$S(t) = e^{-H(t)}.$$

2.3 Competing risks

Sometimes we are interested in multiple possible events for an individual, called **competing risks**. An example would be that a patient can either go into remission or pass away after treatment has taken place. We are interested in the outcome of both events as the occurrence of either will prevent us from observing the other outcome. Suppose we have K competing risks. Then we observe the minimal failure time from any of these risks, $Y = \text{Min}(X_i, ..., X_K)$ with X_i the time to occurrence of the $i-$th risk. The **cause-specific hazard rate** for risk i is then given by:

$$h_i(t) = \lim_{\Delta t \to 0} \frac{\mathbb{P}(t \leq Y < t + \Delta t, \delta = i | Y > t)}{\Delta t}$$

with $\delta \in \{1, ..., K\}$ indicating the cause of failure. The interpretation of this quantity is similar to above, but now restricted to one specific cause of failure. The overall hazard rate is given by:

$$h_Y(t) = \sum_{i=1}^{K} h_i(t).$$

Similarly to above the cause specific cumulative hazard function is given by:

$$H_i(t) = \int_0^t h_i(s)ds$$

and the survival function is given by:

$$S(t) = e^{-\sum_{i=1}^{K} H_i(t)}.$$

Finally, we write the **cumulative incidence function** F_i as:

$$F_i(t) = \mathbb{P}(Y \leq t, \delta = i) = \int_0^t h_i(s) e^{-H_Y(s)} ds$$

where $H_Y(t) = \sum_{i=1}^{K} \int_0^t h_i(s)ds$. The cumulative incidence function is not a real distribution function as it does not necessarily converge to 1 but instead converges to the probability of observing cause of failure i, it can therefore be interpreted as the distribution function associated with cause of failure i when there are multiple possible causes of failure.

2.4 Censoring

Something which is often present in survival data is that observations can be **censored**. This happens when we know the survival time of individuals only when it has taken place within a certain interval. Often this occurs in medical trials where the follow-up will only span a fixed time period, say 4 years. In that case we can only know the survival time for individuals who have started treatment within this period if they have had an event within 4 years after starting the study. This type of censoring is known as **right censoring**. Similarly we may know that an individual has had an event within the period of the study, but not know when treatment first began. This would be called **left censoring**. In this thesis we will focus solely on right censoring.

In right-censored data each individual has two quantities of interest; the lifetime X and some censoring time C_r. The observed lifetime is given by $Y = \min(X, C_r)$. In this case the observed lifetime Y will only be equal to the exact lifetime X when $X \leq C_r$. In this thesis we will consider censoring in the context of clinical procedures. This means that every individual is censored at the end of the clinical follow-up period or due to some personal/medical reason, such as emigration or drop-out.

2.5 Likelihood construction

For the techniques considered in this thesis it will be of great importance to construct the likelihood function for survival data. As we do not have full information about all the lifetimes, the likelihood is slightly different than usual. A very important assumption for the likelihood function in this case is that the lifetimes and censoring times of individuals are independent. When we observe the lifetime X_i of an individual i, his/her component in the likelihood function will be given by $f(x_i)$ as usual, where f is the density function of the survival times. When an individual i has their observed lifetime right censored and we observe $C_{r,i}$; the component in the likelihood will be $S(C_{r,i})$ as we know that the real lifetime is larger than $C_{r,i}$. This way we can determine the likelihood by combining these components:

$$L \propto \prod_{i \in D} f(x_i) \prod_{i \in R} S(C_{r,i}) \tag{2}$$

where D is the set of individuals who have had their true lifetimes observed and R is the set whose observation was right-censored.

2.6 Regression models

In real life (medical) data each individual i often has a vector of characteristics Z_i (called covariates or explanatory variables). Some of these covariates may have an association with the survival time of said individual. An example of this would be that smokers are more prone to developing lung problems, therefore a person who smokes will most likely have a lower survival time when we consider lung illness as outcome. A way to account for this is by including these covariates into the model used to describe the survival times. We will discuss one widely used model in this section.

2.6.1 Cox proportional hazards

This semi-parametric model as introduced by Cox [13] is given by defining the hazard rate in terms of the baseline hazard rate $h_0(t)$:

$$h(t|Z) = h_0(t)e^{\beta Z}.$$

Note that the baseline hazard rate can be parametric as well as non-parametric. The most important property of this model is that the hazard rates for two individuals are proportional at all times:

$$\frac{h(t|Z_1)}{h(t|Z_2)} = \frac{h_0(t)e^{\beta Z_1}}{h_0(t)e^{\beta Z_2}} = \frac{e^{\beta Z_1}}{e^{\beta Z_2}}$$

which is quite a strong assumption. From this we have that:

$$S(t|z) = S_0(t)^{e^{\beta Z}}.$$

At times we will be interested in estimating the baseline (cumulative) hazard rates. Tsiatis [14] has shown asymptotic normality and some other properties for the estimator of the cumulative hazard function as proposed by Breslow [15], which is given by:

$$\hat{H}_0(t) = \sum_{i \in D(t)} \frac{1}{\sum_{j \in R(t_i)} e^{\hat{\beta} Z_j}}$$

with $\hat{\beta}$ the MLE estimate of the regression coefficients, $D(t)$ the set of individuals with failures before time t and $R(t_i)$ the set of individuals at risk at time t_i (with t_i the failure time of patient i in $D(t)$). This estimator approximately maximizes the partial likelihood function, but is not very accurate when tied failure times are present. For this reason, Efron [16] suggested using a similar estimator, which better approximates the solution to the likelihood equation when ties are present, but is in turn more computationally intensive. In practice, when the number of tied observations is small, there is little difference between the two. In our further research we will use the R package Survival [17], which uses the estimator proposed by Efron.

2.7 Generating survival times

In our research we will at some point wish to generate survival time outcomes under some (parametric) distribution. While for standard distributions there are many statistical packages available which can generate outcomes for you, for the regression models above this is not quite so evident. A method to generate survival outcomes according to the Cox proportional hazards model is described in section 15.1.

2.8 Counting processes

In this section we will briefly discuss an approach first developed by Aalen [18], because it will be the key to constructing the CUSUM chart for survival outcomes.

As before, let $Y_i = \min(X_i, C_{r,i})$ be the time of observed failure or right-censoring, where we consider a single cause of failure. We denote by $N(t)$ a **counting process**, which has the following properties:

- $N(0) = 0$,
- $N(t) < \infty$ with probability 1,
- $N(t)$ is right-continuous and piecewise constant with jumps of $+1$.

The processes $N_i(t) = I(Y_i \leq t, \delta_i = 1)$ for individual i indicating whether said individual has failed are counting processes ($\delta_i = 1$ indicates uncensored observation). Then the process $N(t) = \sum_{i=1}^{n} N_i(t)$ is also a counting process, measuring the amount of individuals who have experienced death at or before time t. When considering such a process we have some information available to us about what

has happened to our individuals of interest up to time t. This knowledge can be expressed in the pairs (Y_i, δ_i), which tells us the exact time of outcome when $Y_i \leq t$ and knowledge that $Y_i > t$ for the individuals with no observed outcome until that time point. We will denote this history at time (just prior to) t by F_{t^-}. Now, we have using (1) that:

$$\begin{aligned}&\mathbb{P}(t \leq Y_i \leq t + dt, \delta_i = 1 | F_{t^-}) \\ &= \begin{cases} \mathbb{P}(t \leq X_i \leq t + dt, \delta_i = 1 | X_i \geq t, C_i \geq t) = h_i(t)dt =: dH_i(t), & \text{if } Y_i \geq t \\ 0, & \text{if } Y_i < t \end{cases}\end{aligned}$$

Then define $dN(t)$ as the change in the process $N(t)$ in an infinitesimal period $[t, t + dt]$ and consider $Y(t)$ as the number of individuals with observed time $Y_i \geq t$. Correspondingly, let $Y_i(t)$ indicate whether person i has its observed time after t. Then we have:

$$\begin{aligned}\mathbb{E}[dN(t)|T_i \geq t, C_i \geq t] &= \mathbb{E}[\text{Number of observations with } t \leq X_i \leq t + dt, C_i > t + dt | F_{t^-}] \\ &= Y(t)h(t)dt.\end{aligned}$$

Then we denote the **intensity process** $\lambda(t)$ as:

$$\lambda(t) = Y(t)h(t). \tag{3}$$

Finally the **cumulative intensity process** $\Lambda(t)$ is written as:

$$\Lambda(t) = \int_0^t \lambda(s)ds. \tag{4}$$

In this case $dN_i(t)$ has an approximate Bernoulli distribution with probability $d\Lambda_i(t) = h_i(t)dt$ of observing an unfavourable outcome $dN_i(t) = 1$ when $Y_i \geq t$.

According to Klein & Moeschberger [12] section 3.6 the contribution to the likelihood of an individual i for the approximate Bernoulli process introduced above is proportional to:

$$d\Lambda_i(t)^{dN_i(t)}(1 - d\Lambda_i(t))^{1-dN_i(t)}. \tag{5}$$

Integrating this quantity over the range $[0, \tau]$ would then give a contribution to the likelihood of $\lambda_i(t)^{dN_i(t)} e^{-\int_0^\tau d\Lambda_i(s)ds}$. Then the full likelihood for n observations up to time t would be given by:

$$L = \left(\prod_{i=1}^n \lambda_i(t)^{dN_i(t)} \right) e^{-\sum_{i=1}^n \int_0^t d\Lambda_i(t)}. \tag{6}$$

When considering right-censored data, where $\lambda_i(t) = Y_i(t)h_i(t)$ with $Y_i(t) = 1$ when $t \leq C_i$ and $Y_i(t) = 0$ when $t > C_i$, the contribution to the likelihood is given by:

$$L = \left(\prod_{i=1}^n h_i(t)^{\delta_i} \right) e^{-\sum_{i=1}^n \int_0^t d\Lambda_i(t)} \tag{7}$$

where δ_i is a reverse censoring indicator at time, such that:

$$\delta_i = \begin{cases} 1, & \text{if outcome } i \text{ observed} \\ 0, & \text{if outcome } i \text{ censored} \end{cases}$$

Note how these expressions look like continuous versions of the likelihood as constructed in (2). The power of considering a counting process is in the fact that we are able to consider the likelihood at any time $t > 0$, instead of evaluating the likelihood after observing the outcomes.

2.9 Characteristics of survival distributions

In this section we state some characteristics of distributions which are often used to model survival outcomes and will be used further on in this thesis. We will limit ourselves to the exponential and Weibull distributions. Note that there are different possible parametrisations for these distributions. We list their parametrizations and some key properties used in this thesis in Table 1.

Characteristic	Distribution	
	Exponential	Weibull
Parameter	Scale $\lambda > 0$	Shape $\lambda > 0$ Scale $\theta > 0$
Hazard function	$h_i^\mu(t) = \lambda e^\mu e^{\beta' Z_i}$	$h_i^\mu(t) = \frac{\lambda}{\theta}\left(\frac{t}{\theta}\right)^{\lambda-1} e^\mu e^{\beta' Z_i}$
Cumulative hazard function	$H_i^\mu(t) = \lambda t e^\mu e^{\beta' Z_i}$	$H_i^\mu(t) = \left(\frac{t}{\theta}\right)^{\lambda} e^\mu e^{\beta' Z_i}$
Inverse null cumulative hazard function	$H_0^{-1}(t) = \frac{t}{\lambda}$	$H_0^{-1}(t) = \theta t^{\frac{1}{\lambda}}$
Survival function	$S_i^\mu(t) = e^{-\lambda t e^\mu e^{\beta' Z_i}}$	$S_i^\mu(t) = e^{\left(\frac{t}{\theta}\right)^{\lambda} e^\mu e^{\beta' Z_i}}$
Density function	$f_i^\mu(t) = \lambda e^\mu e^{\beta' Z_i} e^{-\lambda t e^\mu e^{\beta' Z_i}}$	$f_i^\mu(t) = \frac{\lambda}{\theta}\left(\frac{t}{\theta}\right)^{\lambda-1} e^\mu e^{\beta' Z_i} e^{\left(\frac{t}{\theta}\right)^{\lambda} e^\mu e^{\beta' Z_i}}$

Table 1: Some distributions with their characteristic functions, including risk adjustment as well as a deviated hazard by a factor of e^μ. It is possible to recover the non risk-adjusted as well as non deviated quantities by setting $e^{\beta' Z_i} = 1$ and/or $e^\mu = 1$ respectively (which we call the baseline characteristic). The subscript indicates an individual, while the superscript indicates the factor e^μ by which the hazard deviates from the baseline hazard. When the subscript is 0, we indicate the baseline hazard. This notation is consistent as there is no zeroth person.

3 Renewal processes and queueing theory

Some of the proofs later on in this thesis require some basic results from operations research. Therefore, this section will summarise the basic knowledge necessary to understand those proofs. The basis for this theory was taken from the lecture notes written by Kallenberg [19]. This chapter will be written assuming that we are interested in a medical application of this theory.

3.1 Poisson processes

In operations research we are sometimes interested in a process where new parts (patients) arrive according to some random mechanism. To model these arrivals we often assume that the inter arrival times, which are the times between two consecutive arrivals, are exponentially distributed with some parameter λ. The exponential distribution is often used because it is the only continuous distribution which has the beautiful property of memorylessness (see chapter 2.2 of Rice [20]). Mathematically this property is stated as follows:

Let X be a continuous random variable which has non-negative outcomes. Then the distribution of X is **memoryless** if for any non-negative real numbers t and u we have that:

$$\mathbb{P}\left(X > t + u | X > t\right) = \mathbb{P}\left(X > u\right).$$

This property implies that no matter how long you have been waiting for a new patient to arrive, the probability of a new patient arriving after some fixed time u from now will remain the same.

First we start with some renewal theory. Let X_i be the $i-$th **inter arrival time** for $i \in \mathbb{N}_{>0}$ and assume that the inter arrival times are independent and identically distributed. Now define $S_0 = 0$ and $S_n = \sum_{i=1}^{n} X_i$ as the $n-$th **arrival time**. Now define:

$$N(t) = \sup\{n | S_n \leq t\},\ t \geq 0$$

as the total amount of patients which have arrived at or before time t, which is known to be a **renewal process**.

We will only be interested in the case where the inter arrival times are exponentially distributed with parameter λ. In this case $N(t)$ is called a **Poisson process**, sometimes also called a Poisson point process. This is because we then have that:

$$\mathbb{P}\left(N(t) = n\right) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}$$

which is the density of the Poisson distribution with parameter λt. Some interesting properties of a Poisson process are listed below (without proof, see section 1.1 of Kallenberg [19]). Let $t, s \geq 0$:

- $N(t+s) - N(t)$ has the same distribution as $N(s)$,
- $N(t+s) - N(s)$ and $N(s)$ are independent,
- $\mathbb{E}\left[N(t)\right] = \lambda t$.

Using the definitions above it is easy to see that:

$$\mathbb{P}(S_n > t) = \sum_{k=0}^{n-1} \mathbb{P}\left(N(t) = k\right) = \sum_{k=0}^{n-1} \frac{e^{-\lambda t}(\lambda t)^k}{k!}.$$

Therefore:

$$\mathbb{P}(S_n \leq t) = \sum_{k=n}^{\infty} \frac{e^{-\lambda t}(\lambda t)^k}{k!}.$$

Differentiating we obtain the density function of S_n:

$$f_{S_n}(t) = \frac{1}{(n-1)!}\lambda^n t^{n-1} e^{-\lambda t}$$

which has the density of the gamma distribution with parameters λ and n. The maximum likelihood estimator of λ in the Poisson process above at a fixed time t is given by:

$$\hat{\lambda} = \frac{N(t)}{t}. \tag{8}$$

3.2 Queueing theory

A prominent part of operations research is comprised of queueing theory. In this field we consider queues where (in our context) patients arrive to the hospital to be treated. In queueing theory we often assume that arrivals happen according to a Poisson process, and that patients are treated according to an exponential distribution, meaning that they will leave the hospital after an exponentially distributed treatment time. Additionally, it might be that the hospital only has a certain amount of doctors present, which means that some patients will have to wait before starting their treatment. The queue described above would then be denoted as a M/M/c queue, where M stands for Markovian or memoryless Poisson arrivals and exponentially distributed treatment times and c stands for the amount of doctors present to treat patients.

3.2.1 M/G/∞ queue

Consider a M/M/∞ queue (infinite doctors), but instead of exponentially distributed treatment times we consider general treatment times (any distribution). This is then noted as an M/G/∞ queue. Consider the following proposition from Auria [21] (which is given here without proof):

Proposition 3.2.1. *Given an M/G/∞ queue where the arrival process has rate λ and the service times are independent and have common distribution G. Then at a fixed time $t \geq 0$ the number of patients in the system ($Q(t)$) and the number of patients that left the system ($D(t)$) in this interval are independent and distributed according to a Poisson distribution with parameters respectively $\lambda^Q(t) = \frac{\lambda}{t}\int_0^t \overline{G}(t-s)ds$ and $\lambda^D(t) = \frac{\lambda}{t}\int_0^t G(t-s)ds$. Additionally, the departure process $D(t)$ is a Poisson process.*

This proposition will be used in section 5.4 as well as in section 6.4.2 to determine the distribution of the departure process in hospitals, as this quantity is of great interest for our main result.

4 Likelihood ratio tests

By the Neyman-Pearson lemma we know that likelihood ratio tests are the most powerful tests for simple hypotheses, and in some cases they can also be uniformly most powerful for composite hypotheses. Therefore likelihood ratio tests are often a good choice for a test statistic, if the problem allows it. In this section we will introduce some tests which are based on the likelihood ratio. This theory sets the foundation for the method developed by Biswas & Kalbfleisch [2], which is stated in section 5. Inspired by their result we introduce a new method for survival data in section 6, which is the main result of this thesis.

This section focusses on tests using discrete time outcomes, focussing especially on binary outcomes. As the goal of this thesis is to apply the methods on a survival data set from the Dutch National Arthoplasty Register (LROI 12), this means that the data would need to be dichotomized before these methods can be used, which results in a loss of (valuable) information. In sections 5 and 6 however we consider methods which were specifically designed for survival outcomes and therefore use more complete information to arrive at conclusions.

We start this section by introducing the sequential probability ratio test (SPRT), which is used to test a simple hypothesis of a change in parameter. Then we consider the discrete time CUSUM chart in section 4.2, an extension of the SPRT, which is a likelihood ratio test testing whether the quality of care at an insitution has decreased by a **fixed** factor. Later on in section 4.3 we consider a generalization of the discrete time CUSUM chart called the (MAX)GLR chart, which tests whether the quality of care at an institution has decreased by **some** unknown factor. This means that these charts should be used in different scenarios. The CUSUM chart when we have some fixed level of quality reduction which we want to detect, and the (MAX)GLR chart when there is no a-priori information or required detection level. Afterwards, we state two optimality properties of the CUSUM chart which were shown by Lorden [22] and Moustakides [23]. Finally, we introduce the Generalized Likelihood Ratio, which has already been explored by both Wald [5] and Lorden [22]. Continuous time generalizations of these charts are then introduced in sections 5 and 6, called the CTCUSUM and CT(MAX)GLR charts respectively.

The structure of this section was inspired by the thesis written by Kok [9], whereas the notation was taken from chapter 6 of Siegmund [24].

4.1 Sequential probability ratio test

First, we will discuss the Sequential Probability Ratio Test (SPRT) as introduced by Wald [5]. Suppose we have a sequence of outcomes $Y_1, Y_2, ...$ and we want to determine whether these outcomes come from the distribution with density f_{θ_0} or f_{θ_1}. The SPRT is a sequential likelihood ratio test which tests these point hypothesis of the form $H_0 : f_\theta = f_{\theta_0}$ versus $H_1 : f_\theta = f_{\theta_1}$. This test is unique in the way that it tests the hypothesis after every outcome has been observed, which is why it is called a sequential test. The test statistic after observing n outcomes is given by:

$$S_{SPRT,n} = \sum_{i=1}^{n} \ln\left(\frac{f_{\theta_1}(Y_i)}{f_{\theta_0}(Y_i)}\right) = S_{SPRT,n-1} + \ln\left(\frac{f_{\theta_1}(Y_n)}{f_{\theta_0}(Y_n)}\right) =: S_{SPRT,n-1} + W_n.$$

After each observation, the test statistic is determined and one of the following 3 choices are made:

- Reject the null-hypothesis if $S_{SPRT,n} \geq \ln(B)$;
- Accept the null-hypothesis if $S_{SPRT,n} \leq \ln(A)$;

- Observe an additional outcome if none of the above applies;

where A and B are some positive constants. Wald [5] has shown that if we wish for the probability after n observations of a type 1 error to be α and the probability of a type 2 error to be β, then for all practical purposes it is sufficient to take $A = \frac{1-\beta}{\alpha}$ and $B = \frac{\beta}{1-\alpha}$, as then the true type 1 and 2 probabilities α' and β' will approximately adhere to $\alpha' \leq \alpha$ and $\beta' \leq \beta$ (especially when α and β are small). Additionally, Wald has shown that the values above minimize the amount of observations necessary to rightfully reject or accept the null-hypothesis using the SPRT, when the desired probability of a type 1 and type 2 error are α and β respectively.

4.2 Cumulative sum chart

Suppose we again have a sequence of outcomes $Y_1, Y_2, ...$ but this time we are not interested in determining which distribution they come from, but in signalling a change in the distribution at some point in time. For a more precise definition we use the book written by Siegmund [24]. Define P_ν as the probability that $Y_1, ..., Y_{\nu-1} \sim f_{\theta_0}$ and $Y_\nu, ... \sim f_{\theta_1}$ with $\nu = 1, 2, ...$, so that P_ν is the probability that the change in distribution happens at time ν. When this change happens we call the process **out of control**. Similarly, let P_0 be the probability that the change does not happen at all, which in turn means the process stays **in control**. We would like to stop the cumulative sum chart at a point τ, called a **stopping time**, such that the difference between τ and ν is as small as possible. However, we would like the chart not to stop at all if the process stays in control. This is not possible as a false signal can always occur, therefore we would like to make the in control value of τ as big as possible. Mathematically, we would thus like to solve the following problem:

$$\begin{aligned} &\text{minimize} && \sup_{\nu \geq 1} \mathbb{E}_\nu(\tau - \nu + 1 | \tau \geq \nu)) \\ &\text{subject to} && \mathbb{E}_0[\tau] \geq B \end{aligned} \tag{9}$$

for some large constant B.

This problem can be solved using a **cumulative sum chart (CUSUM)** as introduced by Page [3]. Consider, after having observed n outcomes, the null hypothesis $H_0 : Y_1, ..., Y_n \sim f_{\theta_0}$ and the alternative hypothesis $H_1 : Y_1, ..., Y_{\nu-1} \sim f_{\theta_0}$ and $Y_\nu, ..., Y_n \sim f_{\theta_1}$ for $1 \leq \nu \leq n$. The likelihood ratio test for testing whether at least one of the H_ν is true against H_0 is then given by the CUSUM defined as:

$$S_n = \max_{1 \leq k \leq n} \sum_{i=k}^{n} \ln\left(\frac{f_{\theta_1}(Y_i)}{f_{\theta_0}(Y_i)}\right) = \sum_{i=1}^{n} \ln\left(\frac{f_{\theta_1}(Y_i)}{f_{\theta_0}(Y_i)}\right) - \min_{1 \leq k \leq n} \sum_{i=1}^{k} \ln\left(\frac{f_{\theta_1}(Y_i)}{f_{\theta_0}(Y_i)}\right).$$

Alternatively, it is possible to rewrite this using $W_n = \ln\left(\frac{f_{\theta_1}(Y_n)}{f_{\theta_0}(Y_n)}\right)$ to obtain:

$$S_n = \max\left(0, S_{n-1} + W_n\right). \tag{10}$$

Note that this means that the chart resets when it reaches the value 0, as the chart can no longer decrease further. For this reason the CUSUM chart can not build up negative credits so that detection speed is not impeded by periods of low failure rates. Page [3] proposed to reject the null-hypothesis when the chart would reach a certain predefined value h called the **control limit**. Mathematically, this can be stated in terms of a stopping time τ:

$$\tau = \inf\left\{n : S_n \geq h\right\}.$$

Note that this chart is really just a series of SPRTs with lower threshold 0 and upper threshold h where as soon as we accept H_0 we restart the chart at zero, only stopping the chart when we reject H_0 at the upper threshold h. Two important quantities when discussing CUSUM charts are the in control and out of control **average run lengths**. These quantities are defined as follows:

$$ARL_0 = \mathbb{E}_0[\tau] \tag{11}$$
$$ARL_1 = \mathbb{E}_1[\tau] \tag{12}$$

which are the average times it takes for the chart to reject the null-hypothesis when the null-hypothesis is true (in control) or when the alternative hypothesis is true (out of control). Sometimes it is more convenient to consider the **median run lengths**, which are then defined as the median of the distribution of τ.

Let $N_1 = \inf\{n : \sum_{i=1}^{n} W_i \geq h\}$ the time it takes for one SPRT to stop. It then follows from a simple application of Wald's equation (see [24]) that:

$$\mathbb{E}_i[\tau] = \frac{\mathbb{E}_i[N_1]}{\mathbb{P}_i(\sum_{i=1}^{N_1} W_i \geq h)}$$

for $i = 0, 1$, which are the expected run lengths of one SPRT to stop under the null and alternative hypothesis respectively, divided by the probability of rejecting the null-hypothesis in the time it takes for one SPRT to stop.

4.2.1 Optimality properties

In this section we will state the asymptotic optimality properties which have been shown for the CUSUM chart by Lorden [22].

Adhering to the notation and assumptions in section 4.2, our goal is to minimize the conditional expectation of $\tau - \nu$ where τ is the stopping time and ν the time of change of parameter. Let E_ν denote the expectation under P_ν and define:

$$\overline{\mathbb{E}}_1[N] = \sup_{\nu \geq 1} \mathbb{E}_\nu[(\tau - \nu + 1)|Y_1, ..., Y_{\nu-1}] \tag{13}$$

which represents the worst expected detection delay, and is in turn the quantity we wish to minimize. Conversely, we would like for the stopping time to be offset by the requirement that the amount of "false alarms" should be controlled by means of a condition of the form $E_0[N] \geq \gamma$ for some pre-defined $\gamma > 0$.

Lorden [22] has shown that for members of the **Exponential family**, the CUSUM is optimal in an asymptotic sense, such that the worst expected detection delay in equation (13) is minimized. Besides this, he discusses how to determine the optimal threshold value h. Unfortunately, this value depends on the true value of the alternative hypothesis parameter θ_1. Even though the true value of θ_1 is not known in most practical applications, the procedure to determine h is still useful when an approximate value of θ_1 is available. Besides this, it is argued that choosing $h \approx |\ln(\alpha)|$ with α the required type 1 error rate is sufficient for most applications. Later on, Moustakides [23] proved that the CUSUM stopping time is optimal in a more general (non-asymptotic) sense, where optimality was defined similar to Lorden.

4.3 Generalized likelihood ratio

Both the SPRT and CUSUM chart require us to define a simple alternative hypothesis H_1 of the form $\theta_1 = K$ for some constant K. In our medical application, this would mean that we specify the rate

at which we expect failures to happen under the alternative in advance. If we choose this constant incorrectly, we are much more likely to draw the wrong conclusion. Suppose that a change in the distribution of interest does happen somewhere along our observations, but this change is quite minor. If we then choose the alternative hypothesis wrongly we could detect this change when it is far too late. Additionally, it is often not clear in advance which values for the alternative hypothesis are of interest as there is no historical data to see which changes are relevant to detect. Finally, when the parameter of interest changes at some time point τ, there is no guarantee that it will stay at this new value, as the rate of failure can change continuously in most practical applications.

To solve this problem, we can use a **generalized likelihood ratio** test, which tests composite hypotheses. Suppose again that we observe $Y_1, Y_2, ...$ and we would like to determine whether these observations come from the distribution f_{θ_0} or not. In that case we test the null-hypothesis $H_0 : \theta = \theta_0$ against $H_1 : \theta \in \Theta \setminus \{\theta_0\}$. The test statistic is then given by:

$$S_{GLR,n} = \sup_{\theta} \sum_{i=1}^{n} \ln\left(\frac{f_{\theta}(Y_i)}{f_{\theta_0}(Y_i)}\right) = \sum_{i=1}^{n} \ln\left(\frac{f_{\widehat{\theta}_n}(Y_i)}{f_{\theta_0}(Y_i)}\right) \tag{14}$$

with $\widehat{\theta}_n$ the maximum likelihood estimate of θ using the first n observations. Again, we reject the null-hypothesis after n observations if $S_{GLR,n} \geq h$ for some pre defined value of h. This GLR corresponds to the above SPRT when we do not have a simple alternative hypothesis.

Consequently, the GLR alternative for the CUSUM tests hypotheses of the form $H_0 : Y_1, Y_2, ... \sim f_{\theta_0}$ versus $H_1 : Y_1, ..., Y_{\nu-1} \sim f_{\theta_0}, Y_{\nu}, ... \sim f_{\theta}, \theta \in \Theta \setminus \{\theta_0\}$. Adopting the notation used by Kok [9] we will call this test the **maxGLR**. The test statistic is then given by:

$$S_{maxGLR,n} = \max_{1 \leq k \leq n} \sum_{i=k}^{n} \ln\left(\frac{f_{\widehat{\theta}_{n-k}}(Y_i)}{f_{\theta_0}(Y_i)}\right) \tag{15}$$

where $\hat{\theta}_{n-k}$ is the ML estimate of θ over the last $n-k$ observations. The null-hypothesis is rejected when the chart surpasses a certain threshold h. Note that both the GLR and maxGLR chart require the information for all data points up to time n, therefore there is no recursive way to determine them. This makes it more computationally extensive to evaluate these charts.

The differences between the GLR and maxGLR charts are that the GLR chart always uses all available observations to calculate the ML estimate of θ, while the maxGLR uses the MLE over the last k observations, where $1 \leq k \leq n$. Besides this, the GLR chart uses all available observations to construct the chart while the maxGLR chart only uses the last $n-k$ observations to determine the chart, where k as above is chosen to maximize the value of the chart. Note that when the change in distribution does not happen immediately, the ML estimate used in the GLR chart may build up negative credits, depending on the specified model. This will be one of the key factors in section 6 why we will prefer to use the maxGLR over the GLR, even in a continuous time setting where we consider real-time outcomes.

4.3.1 Optimality properties

Lorden [25] has shown that asymptotic optimality properties similar to the ones discussed in section 4.2.1 can also be derived for the generalized likelihood ratio chart. Kok [9] interprets this as follows: Suppose the true value of the alternative hypothesis parameter θ_1 is known and the in control average run length is taken equal for the CUSUM and GLR. Then the average run length under the alternative hypothesis for the GLR was greater than that of the CUSUM by some factor proportional to the in control run length divided by the Kullback-Leibler divergence of the two distributions, as the in control run length goes to infinity. This means that not knowing the out of control parameter comes at a cost

to the run length under the alternative. In most situations however, the real out of control parameter will not be known a-priori, possibly causing worse detection times using a CUSUM chart. In section 7.3 we will study this property for a continuous time version of the CUSUM chart.

4.4 Risk-adjustment

In the problems above we had assumed that the outcome variables $Y_1, Y_2, ...$ were identically distributed in accordance to some distribution function f_θ, specified by the parameter θ. In practical applications however it is often necessary to adjust the parameters of interest based on the **risk** of an individual. In a medical context this risk can be seen as a factor which increases the probability of experiencing an unfavourable outcome, for example the extra risk of contracting lung problems by smoking. It is possible to include these factors into the model in the following way. Suppose for every individual outcome variable Y_i we have an associated vector of **covariates** Z_i (as introduced in chapter 2.6). We would then like for some parameter of the distribution to be dependent on this vector of covariates, so that $Y_i \sim f_{\theta_i}$. There are multiple ways to model this dependence, an example often used for the Bernoulli distribution in CUSUM charts is by letting the probability of success θ_i for person i depend on the covariates of person i by means of logistic regression:

$$\log\left(\frac{\theta_i}{1-\theta_i}\right) = \beta' Z_i$$

with β a vector of coefficients of the same length as Z_i. The CUSUM procedure (in the case of Bernoulli distributed outcomes) then tests the hypotheses $H_0 : OR = 1$ versus $H_1 : OR = R_A > 1$ for some pre-defined value of R_A, where OR is the **odds ratio**. Depending on the specified distribution function this will of course differ, as we will see in chapter 5. The risk-adjustment can also be applied to the GLR in a similar manner.

5 Continuous time CUSUM

This section introduces the method developed by Biswas & Kalbfleisch [2], which cleverly combines the theory about Counting Processes (see section 2.8) with the theory about discrete time CUSUM charts (see section 4.2). By doing so they create a continuous time CUSUM (CTCUSUM) chart, which can be evaluated at every time point instead of limiting ourselves to evaluating the chart at fixed times post treatment. Begun et al. [8] have arrived at a similar model using a different approach, and applied it to a data set from the National Joint Registry in the UK. The main result of this thesis is a generalisation of this chart called the CTGLR chart and is constructed in section 6.

Although the discrete time CUSUM chart was not developed with survival outcomes in mind, in this section we introduce the method of Biswas & Kalbfleisch [2] based on a CUSUM procedure which does incorporate survival outcomes. In this section we consider the time to revision of patients who have had primary hip transplant surgery as the outcome of interest. One option is to consider as outcome the (binary) status of a patient (1 for dead and 0 for alive) after a certain amount of time has passed post surgery. In this way, the outcome can be modelled using a Bernoulli distribution. This is necessary as the discrete time CUSUM defined in section 4.2 can only be evaluated at the (time of) outcomes. To give an example, van Schie et al. [11] construct a Bernoulli CUSUM chart using revision one year post transplant as outcome. In this section, we will introduce the information provided by the survival of a patient at every time point into the model in the hope of improving the detection speed of the chart.

Similarly to the discrete time CUSUM chart defined in section 4.2, the CTCUSUM chart will test a **point hypothesis** of a **fixed** decrease in quality of care at some institution. A generalization of this chart, similar to the (max)GLR chart considered in section 4.3, will be introduced in section 6, which tests a **composite hypothesis** of a decrease in the quality of care by some **unknown** factor. The latter is a more general test requiring less prior information about the problem, but we will see that this generalization comes at a certain price.

5.1 Notation

We will adhere as much as possible to the notation used in Biswas & Kalbfleisch [2]. Let X_i denote the time from the primary procedure to the time of revision for patient i. Then define S_i as the time of primary procedure from some starting point of the study. Then the chronological time of failure is given by $T_i = S_i + X_i$. Let $C_{r,i}$ be the time to right-censoring from primary procedure for patient i, and C_i the chronological time of censoring, then $C_i = S_i + C_{r,i}$. For every patient we also have covariates which are denoted by the $p-$vector Z_i. The covariates for all patients can then be combined row-wise to form the matrix Z. Additionally, we assume that there is a known (risk-adjusted) null-distribution for X_i, defined by the hazard function $h_i(x)$ (note that x here is the time to revision). For $\mu > 0$ define $h_i^\mu(x) = h_i(x) \cdot e^\mu$. Let this notation also carry over to the cumulative hazard rate $H_i^\mu(x)$ as well as the survival function $S_i^\mu(x)$ and the density function $f_i^\mu(x)$. With this notation the subscript indicates the risk-adjustment term for person i and the superscript denotes the factor e^θ by which the hazard differs from the null-rate (which is indicated without superscript). A characteristic without superscript indicates that $\mu = 0$, so that the notation looks neater. Additionally, we denote by $h_0(x)$ the non risk-adjusted hazard rate under the null hypothesis (there is no 0-th person making this notation consistent).

Now define $N^A(t) = \sum_{i \geq 1} \mathbb{1}_{\{S_i \leq t\}}$ to be the number of primary procedures (transplants) in $[0, t]$. Define $\tilde{N}_i^D(t) = \mathbb{1}_{\{T_i \leq t\}}$ as a failure indicator for patient i up to time t. Choose $C > 0$ and define

$$Y_i(t) = \mathbb{1}_{\{S_i \leq t \leq S_i + C \cap T_i \cap C_i\}}$$

an indicator whether patient i is **active**. If a person is not active, we call them **inactive**. This means that people are only active after they have had a primary procedure and only up until the point that they have either failed, been censored or reached C years post transplant. Define $N_i^D(t) = \int_0^t Y_i(u)d\tilde{N}_i^D(u)$ for $t > 0$ as the counting process for a qualifying failure of patient i. Define the history (filtration) for $N_i^D(t)$ as

$$\mathcal{F}_t = \sigma\left\{N^A(u), N_i^D(u), Y_i(u), Z_i, N^A(t), i = 1, 2, ..., n_t = N^A(t) : 0 \leq u < t\right\}.$$

Finally, define the **cumulative intensity** (see (4)):

$$\Lambda_i^\mu(t) = \int_0^t Y_i(u) \cdot h_i^\mu(u)du \tag{16}$$

with subscript and superscript as above. Note that for indices i which have not yet had a primary procedure (when $t < S_i$) we have that $\Lambda_i^\mu(t) = 0$, due to the definition of $Y_i(t)$. A table with characteristics for some survival distributions used in this thesis can be found in Table 1.

5.2 Cox proportional hazards model

We will consider the CUSUM procedure applied to the Cox proportional hazards model as introduced by Biswas & Kalbfleisch [2], which uses the counting process defined in section 2.8. The only major change is that we will consider outcomes up until C years after the primary procedure, for some $C > 0$, whereas Biswas & Kalbfleisch only consider outcomes up to 1 year post transplant. The Cox regression model based on the chosen covariates is then given by:

$$h_i(x) = h_0(x)e^{Z_i^\top \beta} \text{ for } x > 0.$$

Now using the chronological time t (from the start of the study) define a counting process $N_i^D(t)$ corresponding to the $i-$th patient as above. We then have:

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

Now denote $dH_i(t) = h_i(t-S_i)dt$, and note that this represents the instantaneous hazard of a revision. The term e^μ is the factor by which the hazard at an institution differs from the national rate $dH_i(t)$, or equally the null rate with $\mu = 0$.

We want to calculate a likelihood ratio statistic corresponding to a test of $\mu = 0$ versus $\mu = \theta$ ($\theta > 0$ known), therefore testing whether the quality of transplantations has decreased. Consider the likelihood based on $dN_i^D(t)$, the response in the interval $(t, t + dt]$ conditional on the information available up to time t. We obtain the likelihood ratio in this interval using equation (5):

$$\begin{aligned} LR'_\theta(t, t + dt) &= \frac{\prod_{i\geq 1}\left(e^\theta d\Lambda_i(t)\right)^{dN_i^D(t)}\left(1 - e^\theta d\Lambda_i(t)\right)^{1-dN_i^D(t)}}{\prod_{i\geq 1}\left(e^0 d\Lambda_i(t)\right)^{dN_i^D(t)}\left(1 - e^0 d\Lambda_i(t)\right)^{1-dN_i^D(t)}} \\ &= \prod_{i\geq 1}\frac{\left(e^\theta\right)^{dN_i^D(t)}\left(1 - e^\theta d\Lambda_i(t)\right)^{1-dN_i^D(t)}}{\left(1 - e^0 d\Lambda_i(t)\right)^{1-dN_i^D(t)}}. \end{aligned}$$

Remember that individuals stop providing information to the chart after they are censored, but the information acquired until the time of censoring is taken into regard. The implications of this

construction will be discussed below. Now using a repeated conditioning argument, the likelihood based on the information up to time t can be calculated using equation (6) and is equal to:

$$LR_\theta(t) = \prod_{i\geq 1} \frac{\left(e^\theta\right)^{N_i^D(t)} e^{-e^\theta \Lambda_i(t)}}{e^{-\Lambda_i(t)}}.$$

As a consequence, the log likelihood ratio up to time t is given by:

$$U_\theta(t) := \ln(LR_\theta(t)) = \sum_{i\geq 1} \ln\left(\frac{\left(e^\theta\right)^{N_i^D(t)} e^{-e^\theta \Lambda_i(t)}}{e^{-\Lambda_i(t)}}\right) \tag{17}$$

$$= \sum_{i\geq 1} \theta N_i^D(t) + \left(-e^\theta \Lambda_i(t) + \Lambda_i(t)\right) \tag{18}$$

$$= \theta N^D(t) - (e^\theta - 1)\sum_{i\geq 1} \Lambda_i(t). \tag{19}$$

Because we would like to test the hypothesis of a change of hazard rate starting from some patient τ, the CUSUM chart has a cut-off at zero using a similar argument as in section 4.2. To continuously update the chart we are interested in the increments dUt defined as:

$$dU_\theta(t) = \theta dN^D(t) - (e^\theta - 1)\sum_{i=1}^{n} d\Lambda_i(t). \tag{20}$$

The (one-sided) Continuous time CUSUM chart is then given by:

$$G_\theta(t + dt) = \max(0, G_\theta(t) + dU_\theta(t)) \text{ for } t > 0 \tag{21}$$

which is equivalent to:

$$G_\theta(t) = U_\theta(t) - \min_{0\leq s\leq t} U_\theta(s) \text{ for } t > 0.$$

Note that if we wanted to test whether there was an increase in quality (therefore decrease in the amount of revisions) we could plot:

$$G_\theta^-(t) = \min(0, G_\theta^-(t) - dU_\theta(t)) \text{ for } t > 0.$$

We will not pursue this further in this thesis.

5.3 Definition

We summarise the results from the previous section in the following definition.

Definition 5.3.1. *The **continuous time cumulative sum chart (CTCUSUM)** is given by:*

$$G_\theta(t) = \max\left(0, \theta N^D(t) - (e^\theta - 1)\sum_{i\geq 1} \Lambda_i(t)\right). \tag{22}$$

This chart is used to test the hypotheses of a change in cumulative intensity starting from patient ν:

$$\begin{aligned} H_0 &: X_1, X_2, ... \sim \Lambda_i \\ H_1 &: X_1, ..., X_{\nu-1} \sim \Lambda_i \\ &\quad X_\nu, X_{\nu+1}, ... \sim \Lambda_i^\theta \end{aligned}$$

for $\nu \geq 1$ unknown and $\Lambda_i^\theta = e^\theta \cdot \Lambda_i$ the risk-adjusted baseline cumulative intensity multiplied by e^θ and $\theta > 0$ a constant chosen in advance. The null hypothesis is rejected at time t when $G_\theta(t) \geq h$ for some $h > 0$ chosen in advance, called the ***control limit****.*

Note the similarity of these definitions to those we have seen in section 4.2. Similarly to how the discrete time CUSUM chart was defined in equation (10), we can calculate it recursively. However, we were limited to evaluating the chart after a certain period of time had passed post transplant. Using this model we can calculate the chart at any time $t > 0$. In practice, we can see in the definition of $dU_\theta(t)$ (20) that the chart will only jump up by θ when we observe a revision (notice how censored observations do not cause a jump), and will drift downwards by the sum of the instantaneous hazard of the people at risk of having a revision at that time, multiplied by the *excess risk* $e^\theta - 1$. The censoring mechanism makes it so that observations which are censored before their failure time only contribute to the downward motion of the chart, and can never contribute to an increase. It's also possible to erase the contribution of censored observations from the chart completely, either by simply not considering them (leading to more false positives) or by compensating the negative influence of the patient at the censoring time (leading to a delay in detection). Both methods ignore the available information, as the Bernoulli CUSUM chart would do. In most cases it is therefore preferable to include censored observations into the chart.

An important difference between the discrete and continuous time CUSUM chart is that the continuous time chart trends upwards by θ when we observe an unfavourable outcome, independent of the covariates of the person in question, while in a risk-adjusted discrete time (Bernoulli) CUSUM the chart will jump upward or downward depending on the value of the alternative hypothesis and the predicted probability of unfavourable outcome. In the continuous chart however, the trend downwards does depend on the covariates of the people which are active at that time. It does this through the (cumulative) baseline hazard as can be seen in formula (19). This means that the chart decreases depending on how great the risk of failure is of active people (a high hazard rate implies high risk of failure), multiplied by the excess risk. Intuitively this is what we would like to expect from our chart, as it decreases rapidly when many vulnerable people at risk do not experience a failure, and increases by a fixed value whenever we observe a failure.

5.4 An approximation to the ARL

In their article, Biswas & Kalbfleisch [2] derive an approximation to the ARL of the continuous time CUSUM constructed using the Cox PH model. In this section we will repeat this process so that we can use the techniques for a similar proof later on in section 6. The original authors considered only outcomes 1 year post transplant, while we will consider outcomes C years post transplant, for some $C > 0$. Part of the proof will rely on results stated in the Appendix 15.

If the hypothesized rate holds then, by assumption:

$$\mathbb{E}_{X_i}[dN_i^D(t)|\mathcal{F}_t] = e^\mu d\Lambda_i(t) = e^\mu Y_i(t) h_i(t - S_i)dt.$$

Note that using this notation, $Y_i(t)$ contains information about whether a patient undergoes a qualifying failure and whether the patient has a censored observation. Now let $N^D(t) = \sum_{i \geq 1} N_i^D(t)$ as above and define $A(t) = \sum_{i=1}^n \Lambda_i(t)$. Then the CTCUSUM chart is given by:

$$G_\theta(t) = \theta N^D(t) + (e^\theta - 1)A(t) \tag{23}$$

for some $\theta > 0$.

It is natural to assume that patients arrive according to a homogeneous Poisson process (as defined in section 3.1) with rate $\psi > 0$. To study the process in equilibrium, we assume that arrivals begin

at time $t = -C$, this is sufficient for equilibrium since only failures within C years of transplant are viewed as qualifying failures, and thus at $t = 0$ we will have considered all qualifying failures up to that point. Since we begin at time $t = -C$, $S_i + C$ will have a Gamma distribution with scale ψ and shape i as we saw in section 3.1.

Remember that $A(t) = \sum_{i\geq 1} \Lambda_i(t)$. Censoring mechanisms are not considered in this approximation, therefore redefine $Y_i(u) = \mathbb{1}\{S_i \leq u \leq S_i + C, u \leq T_i\}$. We can recover $S_i + C \leq u + C$ from $S_i \leq u$ and combine this with $u \leq S_i + C$. Additionally, using $T_i = S_i + X_i$ we obtain that $u \leq T_i$ transforms into $u \leq X_i + S_i$ and finally into $u + C - (S_i + C) \leq X_i$. Then using the law of total expectation twice (conditioning first on Z_i and then on S_i):

$$
\begin{aligned}
&\mathbb{E}[dA(u)] \\
&= \mathbb{E}\left[\sum_{i\geq 1} Y_i(u) h_i(u - S_i) du\right] \\
&= e^{-\mu}\mathbb{E}\left[\sum_{i\geq 1} \mathbb{1}\{u \leq S_i + C \leq u + C, X_i \geq u + C - (S_i + C)\} e^{\mu} h_i(u - S_i) du\right] \\
&= e^{-\mu}\sum_{i\geq 1} \mathbb{E}\left[\mathbb{1}\{u \leq S_i + C \leq u + C\}\mathbb{1}\{X_i \geq u + C - (S_i + C)\} e^{\mu} h_i(u - S_i) du\right] \\
&= e^{-\mu}\sum_{i\geq 1} \mathbb{E}_{S_i}\left[\mathbb{E}_{Z_i}\left[\mathbb{E}\left[\mathbb{1}\{u \leq S_i + C \leq u + C\}\mathbb{1}\{X_i \geq u + C - (S_i + C)|Z_i\} e^{\mu} h_i(u - S_i|Z_i) du\right]|S_i\right]\right] \\
&= e^{-\mu}\sum_{i\geq 1} \mathbb{E}_{S_i}\left[\mathbb{1}\{u \leq S_i + C \leq u + C\}\mathbb{E}_{Z_i}\left[\mathbb{E}\left[\mathbb{1}\{X_i \geq u + C - (S_i + C)|Z_i\}\right] e^{\mu} h_i(u - S_i|Z_i) du|S_i\right]\right] \\
&= e^{-\mu}\sum_{i\geq 1} \mathbb{E}_{S_i}\left[\mathbb{1}\{u \leq S_i + C \leq u + C\}\mathbb{E}_{Z_i}\left[\mathbb{P}\left(X_i \geq u + C - (S_i + C)|Z_i\right) h_i^{\mu}(u + C - (S_i + C)|Z_i) du|S_i\right]\right] \\
&= e^{-\mu}\sum_{i\geq 1} \mathbb{E}_{S_i}\left[\mathbb{1}\{u \leq S_i + C \leq u + C\}\mathbb{E}_{Z_i}\left[S_i^{\mu}\left(u + C - (S_i + C)|Z_i\right) h_i^{\mu}(u + C - (S_i + C)|Z_i) du|S_i\right]\right] \\
&= e^{-\mu}\sum_{i\geq 1} \mathbb{E}_{S_i}\left[\mathbb{1}\{u \leq S_i + C \leq u + C\}\mathbb{E}_{Z_i}\left[f_i^{\mu}(u + C - (S_i + C)|Z_i)|S_i\right] du\right] \\
&= e^{-\mu}\sum_{i\geq 1} \int_u^{u+C} \mathbb{E}_{Z_i}\left[f_i^{\mu}(u + C - x)\right] \psi \frac{e^{-\psi x}(\psi x)^{i-1}}{(i-1)!} dx du \\
&= e^{-\mu}\psi \int_u^{u+C} \mathbb{E}_{Z_i}\left[f_i^{\mu}(u + C - x)\right] e^{-\psi x} \sum_{i\geq 1} \frac{(\psi x)^{i-1}}{(i-1)!} dx du.
\end{aligned}
$$

Note that the summation is an infinite one, but due to the definition of $Y_i(t)$ only active individuals will contribute towards the term. We obtain:

$$
\begin{aligned}
\mathbb{E}[dA(u)] &= e^{-\mu}\psi \int_u^{u+C} \mathbb{E}_{Z_i}\left[f_i^{\mu}(u + C - x)\right] du &(24)\\
&= e^{-\mu}\psi \mathbb{E}_{Z_i}\left[F_i^{\mu}(C)\right] du &(25)
\end{aligned}
$$

which follows from Lemma 15.2.1. We define the notation:

$$
\gamma := e^{-\mu}\psi \mathbb{E}_{Z_i}[F_i^{\mu}(C)] > 0 \tag{26}
$$

which is a constant. Then $\mathbb{E}[dA(u)] = \gamma du$. Now consider the hospital as a queue. Primary procedures come in with rate ψ and consider the waiting time to be the time to failure, for which we assume that patients have risk-adjusted distribution F_i^{μ}. In this case the failure process can be described by the departure process as in Proposition 3.2.1. We then obtain that $N^D(t)$ is a Poisson process with rate $e^{\mu}\gamma$. Incorporating this into the chart, we have:

$$U_t = \theta N^D(t) - (e^{\theta} - 1)\gamma t + E_t$$

where using Lemma 15.3.1 and assuming that $\mathbb{E}[A(t)]$ exists we obtain that $E_t = (e^{\theta} - 1)(\gamma t - A(t))$ is a zero-mean process. We make the approximation that $E_t = 0$ so that a theoretical result can be obtained.

Note that the CTCUSUM chart $G_\theta(t)$ can be obtained from U_t by resetting the chart whenever U_t reaches 0 and signalling when the process reaches an upper barrier h. Additionally, we know that once U_t (and similarly the CTCUSUM chart) reaches 0, it will stay at this value until at some point it jumps up again to θ. This is because the chart can only experience fixed upwards jumps of size θ per definition. Note that U_t has stationary increments. This is because $U_t - U_s = \theta(N^D(t) - N^D(s)) - (e^{\theta} - 1)\gamma(t - s)$ and $U_{t-s} = \theta N^D(t - s) - (e^{\theta} - 1)\gamma(t - s)$ using that a Poisson Process enjoys the property that $N^D(t) - N^D(s) = N^D(t - s)$ (see 3.1). This means that we can consider every jump from 0 to θ as a **renewal** (see section 3.1).

Suppose that $U_0 = G_0 = 0$ and let:

$$F_0 = \inf\{t > 0 : G_t = \theta\}.$$

By definition of a Poisson Process, $F_0 \sim Exp(e^{\mu}\gamma)$. Therefore the sequence of events $G_t = \theta$, $G_{t^-} = 0$ is a renewal process delayed by F_0. After observing a renewal the process either returns to zero (followed by another renewal) or it crosses the upper barrier $h > 0$. For the approximation, we look at the process U_t with $U_0 = \theta$ and absorbing barriers 0 and h. Let p_R and $(1 - p_R)$ be the probability of absorption at h and 0 respectively. Furthermore, let:

$$T^{(\theta)} = \inf\{t > 0 : U_t \notin (0, h), U_0 = \theta\}$$

be the time to absorption.

In the original process G_t, we denote by J be the number of renewals including the first so that $J \geq 1$ and notice that J is a stopping time (because once there are no more renewals we know that the chart has been stopped by h). It is apparent that:

$$\mathbb{P}(J = j) = (1 - p_R)^{j-1} p_R.$$

After observing the i-th renewal, the waiting time τ_h for absorption at h is increased by $W_i = T_i^{(\theta)} + (1 - \Delta_i)R_i$, where Δ_i is a binary indicator of absorption at h versus a return to 0. R_i represents the time from recurrence of $G_t = 0$ to the next jump to level θ. Finally, $T_i^{(\theta)}$ is the time until the process exceeds h or returns to 0, whichever occurs first. It is now easy to see that:

$$\mathbb{E}[W_i] = \mathbb{E}[T^{(\theta)}] + (1 - p_R)\mathbb{E}[F_0]$$

since $\mathbb{E}[R_i] = \mathbb{E}[F_0]$. Thus:

$$\tau_h = F_0 + \sum_{i=1}^{J} W_i$$

and since J is a stopping time, Wald's identity yields:

$$\mathbb{E}[\tau_h] = \mathbb{E}[F_0] + \mathbb{E}[J]\left[\mathbb{E}[T^{(\theta)}] + (1-p_R)\mathbb{E}[F_0]\right] \tag{27}$$

$$= \mathbb{E}[F_0] + \frac{\mathbb{E}[T^{(\theta)}] + (1-p_R)\mathbb{E}[F_0]}{p_R} \tag{28}$$

$$= \frac{\mathbb{E}[T^{(\theta)}] + \mathbb{E}[F_0]}{p_R} \tag{29}$$

as $\mathbb{E}[J] = \frac{1}{p_R}$ because J is clearly geometrically distributed with parameter p_R.

By definition of a Poisson process we find that:

$$\mathbb{E}[F_0] = \frac{e^{-\mu}}{\gamma}$$

thus we only need to find the expected value of $T^{(\theta)}$ and p_R.

Consider again the process $\{U_t\}$ with $U_0 = 0$ and with absorbing barriers at $-\theta$ and $h-\theta$, it will be apparent later on why we are interested in this construction. Let

$$f^*(\omega) = \mathbb{E}[e^{-\omega U_t}]$$

be the moment generating function of U_t in the unrestricted process. We find that:

$$\begin{aligned}
f^*(\omega) = \mathbb{E}[e^{-\omega U_t}] &= \mathbb{E}\left[e^{-\omega(\theta N^D(t) - (e^\theta - 1)\gamma t)}\right] = e^{\omega(e^\theta-1)\gamma t}\cdot\mathbb{E}\left[e^{-\omega\theta N^D(t)}\right] \\
&= e^{\omega(e^\theta-1)\gamma t}\sum_{k=0}^{\infty} e^{-\omega\theta k}\frac{(e^\mu\gamma t)^k e^{-(e^\mu\gamma t)}}{k!} \\
&= e^{\omega(e^\theta-1)\gamma t}e^{e^\mu\gamma t}\sum_{k=0}^{\infty}\frac{\left(e^{-\omega\theta}e^\mu\gamma t\right)^k}{k!} \\
&= e^{\omega(e^\theta-1)\gamma t}e^{e^\mu\gamma t}e^{e^{-\omega\theta}e^\mu\gamma t} \\
&= e^{\gamma t\left(\omega(e^\theta-1)+e^\mu\left(e^{-\omega\theta}-1\right)\right)}
\end{aligned}$$

using the fact that $N^D(t)$ is a Poisson Process with rate $e^\mu\gamma$, therefore $N^D(t)$ is distributed as Poisson$(e^\mu\gamma t)$. Now let ω_0 be the solution to:

$$f^*(\omega_0) = 1.$$

Note that ω_0 does not depend on t. This can be seen as we can find ω_0 by equating the term in the exponent above to zero:

$$\omega_0(e^\theta - 1) + e^\mu\left(e^{-\omega_0\theta} - 1\right) = 0$$

which does not depend on t. To continue we need a result from Cox & Miller [26] which states that:

Theorem 5.4.1. *Let $S_m = X_1 + ... + X_m$, where X_j are independent random variables with common mgf $\phi(t)$, which is assumed to exist in a real interval containing $t = 0$. Let the RV n be defined as the smallest integer m for which either $S_m \geq \alpha$ or $S_m \leq -\beta$ with $\alpha, \beta > 0$. Thus n can be regarded as the time to absorption for the random walk S_m with absorbing barriers at α and $-\beta$. Let $S = S_n$ and let $F_m(x) = \mathbb{P}(-\beta < S_k < \alpha$ for $k = 1, 2, ..., m-1$ and $S_m \leq x)$. Then we have that*

$$\mathbb{E}\left[e^{tS}z^n\right] = 1 + (z\phi(t) - 1)F(z,t)$$

where $F(z,t) = \sum_{m=0}^{\infty} z^m \int_{-\beta}^{\alpha} e^{tx} dF_m(x)$. Then Wald's identity follows from this equality by setting $z = (\phi(t))^{-1}$.

Applying Wald's identity from this theorem we then obtain, for any ω:

$$\mathbb{E}\left[e^{-\omega U_{T^{(\theta)}}} f(\omega)^{-T^{(\theta)}}\right] = \mathbb{E}\left[e^{-\omega U_{T^{(\theta)}} - T^{(\theta)}\ln(f^*(\omega))}\right] = 1.$$

Evaluating the expected value at ω_0 yields:

$$\begin{aligned}\mathbb{E}\left[e^{-\omega_0 U_{T^{(\theta)}}} f(\omega_0)^{-T^{(\theta)}}\right] &= \mathbb{P}\left(U_{T^{(\theta)}} = (h-\theta)\right)\cdot\mathbb{E}[e^{-\omega_0(h-\theta)-T^{(\theta)}\ln(f^*(\omega_0))}] + \mathbb{P}\left(U_{T^{(\theta)}} = -\theta\right)\cdot\mathbb{E}[e^{\omega_0\theta - T^{(\theta)}\ln(f^*(\omega_0))}] \\ &= p_R\cdot e^{-\omega_0(h-\theta)} + (1-p_R)e^{\omega_0\theta} \approx 1\end{aligned}$$

where we use the fact that $\ln(f^*(\omega_0)) = 0$ and ignoring overshoot across the boundaries (as our stopping time is not exact due to the nature of the chart). We retrieve:

$$p_R \approx \frac{1-e^{-\omega_0\theta}}{1-e^{-\omega_0 h}}.$$

Now let $\eta dt := \mathbb{E}[dU_t]$. We have that:

$$\begin{aligned}\mathbb{E}[dU_t] &= \mathbb{E}[\theta dN^D(t) - (e^\theta - 1)\gamma dt] \\ &= \theta\mathbb{E}[dN^D(t)] - (e^\theta-1)\gamma dt \\ &= \theta e^\mu\gamma dt - (e^\theta - 1)\gamma dt\end{aligned}$$

thus we define $\eta := (\theta e^\mu - e^\theta + 1)\gamma$. If $\eta \neq 0$, then Wald's identity gives:

$$\eta\mathbb{E}\left[T^{(\theta)}\right] = \mathbb{E}[U_{T^{(\theta)}}]$$

and again ignoring overshoot:

$$\begin{aligned}\mathbb{E}\left[U_{T^{(\theta)}}\right] &\approx (h-\theta)p_R + -\theta(1-p_R) \\ &= hp_R - \theta.\end{aligned}$$

When $|\eta| \to 0$, then $|\omega_0| \to 0$ and $p_R \to \frac{\theta}{h}$.

Now let $\sigma^2 dt = \mathrm{Var}(dU_t)$. We have:

$$\begin{aligned}\mathbb{E}[dU_t^2] &= \mathrm{Var}(dU_t) + (\mathbb{E}[dU_t])^2 \\ &= \sigma^2 dt + \left(\theta e^\mu\gamma dt - (e^\theta - 1)\gamma dt\right)^2 \\ &\overset{\eta=0}{=} \sigma^2 dt.\end{aligned}$$

Thus:

$$\mathbb{E}[U_t^2] \overset{\eta=0}{=} \sigma^2 t$$

therefore using Wald's identity:

$$\mathbb{E}\left[U^2_{T^{(\theta)}}\right] \overset{\eta=0}{=} \sigma^2\mathbb{E}[T^{(\theta)}]$$

and finally (ignoring overshoot):

$$\mathbb{E}\left[U^2_{T^{(\theta)}}\right] \approx (h-\theta)^2 p_R + \theta^2(1-p_R).$$

Finally:

$$\mathrm{Var}(dU_t) \overset{\eta=0}{=} \mathrm{Var}(\theta dN^D(t)) = \theta^2\gamma e^{\mu}dt =: \sigma^2 dt.$$

Summarising above results we have when $\eta = 0$:

$$p_R = \frac{\theta}{h} \qquad \mathbb{E}[T^{(\theta)}] = \frac{\mathbb{E}[U^2_{T^{(\theta)}}]}{\sigma^2} = \frac{(h-\theta)^2 p_R + \theta^2(1-p_R)}{\theta^2\gamma e^{\mu}}$$

and when $\eta \neq 0$:

$$p_R \approx \frac{1-e^{-\omega_0\theta}}{1-e^{-\omega_0 h}} \qquad \mathbb{E}[T^{(\theta)}] = \frac{\mathbb{E}[U_{T^{(\theta)}}]}{\eta} = \frac{(\eta-\theta)p_R - \theta(1-p_R)}{(\theta e^{\mu} - e^{\theta} + 1)\gamma}$$

Now rewriting and substituting into equation (29) we obtain that:

$$\mathbb{E}[\tau_h] = \begin{cases} \frac{h}{\eta} - \frac{e^{-\mu}(e^{\theta}-1)}{\eta}\left(\frac{1-e^{-\omega_0\theta}}{1-e^{-\omega_0 h}}\right), & \eta \neq 0 \\ \frac{h^2 e^{-\mu}}{\theta^2\gamma}, & \eta = 0 \end{cases} \tag{30}$$

therefore allowing us to approximate the ARL of the CTCUSUM using a closed-form expression. We have two distinctions, when $\eta = 0$ and when $\eta \neq 0$. We remind the reader that $\eta = 0$ implies $(\theta e^{\mu} - e^{\theta} + 1)\gamma = 0$ and as $\gamma > 0$ this means that:

$$\mu = \ln\left(\frac{e^{\theta}-1}{\theta}\right)$$

thus quantifying $\eta = 0$ in terms of μ and θ.

Now that we have re-iterated the theory already developed by Biswas & Kalbfleisch [2] we will develop a generalized method based on this theory and evaluate its theoretical properties in the next chapter. This generalization will be constructed in the hope that it will be more applicable to real-life data sets, as it will require less prior knowledge to construct and evaluate.

6 Continuous time Generalized Likelihood Ratio chart

In this section we develop two generalizations of the Continuous Time CUSUM chart by Kalbfleisch & Biswas [2], which was stated in section 5. We call these new methods the Continuous Time (Maximized) Generalized Likelihood Ratio charts (CTGLR and CTMAXGLR). These charts are especially useful when we do not know or do not want to choose a value for the alternative hypothesis as was required in the continuous time CUSUM chart.

We first derive and define the CTGLR chart in Definition 6.3.1. The main theoretical result is stated in section 6.5, which is the asymptotic distribution of the CTGLR chart. Deriving from this main result we determine an approximation to the ARL of the CTGLR chart in section 6.6. Finally, we introduce the Continuous Time MAXGLR chart in definition 6.9.1, as a further generalization of the CTGLR chart. Throughout the section, some properties of the CTGLR chart are shown, which are easily generalized for the CTMAXGLR chart.

Some properties of the CT(MAX)GLR chart shown in this section are the non-negativity (Lemma 6.3.1), asymptotic bounds (Lemma 6.8.1) and the equality of the approximate ARL of the CTGLR and CTMAXGLR charts when the process is out of control during the whole study as well as a means to equate their ARLs when this is not the case (Proposition 6.9.1). Additionally, we derive an approximation to the ARL of the CTCUSUM chart using a different method than Biswas & Kalbfleisch [2] in Corollary 6.10.1 and show that the CTCUSUM chart has a worse asymptotic ARL than the CT(MAX)GLR chart when the process is out of control and the value of θ is chosen incorrectly.

6.1 Notation

The notation will stay unchanged from section 5.1, except that instead of $Y_i(t) = \mathbb{1}_{\{S_i \leq t \leq S_i + C \cap T_i \cap C_i\}}$ we will now define

$$Y_i(t) = \mathbb{1}_{\{S_i \leq t \leq T_i \cap C_i\}}$$

as an indicator whether patient i is **active**. If a person is not active, we call them **inactive**. This means that people are only active after they have had a primary procedure and only up until the point that they have either failed or been censored, meaning we no longer restrict ourselves to C years post transplant. We remind the reader that the **cumulative intensity** is given by:

$$\Lambda_i^{\mu}(t) = \int_0^t Y_i(u) \cdot h_i^{\mu}(u) du \tag{31}$$

with $\Lambda_i(t) := \Lambda_i^0(t)$. Note that for indices i which have not yet had a primary procedure, so when $t < S_i$, we have that $\Lambda_i^{\mu}(t) = 0$, due to the definition of $Y_i(t)$. A table with characteristics for some commonly used survival distributions can be found in Table 1.

6.2 Cox proportional hazards model

We will consider a Generalized Likelihood Ratio chart using a proportional hazards assumption for the outcome distribution, based on the theory in section 5. Similarly, we also consider the counting processes as defined in section 5.2 where we defined a counting process $N_i^D(t)$ corresponding to the $i-$th patient such that:

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

The difference will lie in the alternative hypothesis considered. Let e^{μ} be the true factor by which the hazard at the hospital is higher than the baseline hazard. This time we want to calculate a likelihood ratio statistic corresponding to a test of $\mu = 0$ versus $\mu = \theta$ for some **unknown** $\theta > 0$, again testing whether the quality of transplantations has decreased. For the continuous time CUSUM we assumed θ to be known. The likelihood ratio based on $dN_i^D(t)$, using the information from the response in the interval $(t, t + dt]$ conditional on the information available up to time t is then given by (see equation (5)):

$$\begin{aligned} LR'_{\hat{\theta}_t}(t) = LR'_{GLR}(t) &:= \sup_{\theta>0} \frac{\prod_{i\geq 1} \left(e^{\theta} d\Lambda_i(t)\right)^{dN_i^D(t)} \left(1 - e^{\theta} d\Lambda_i(t)\right)^{1-dN_i^D(t)}}{\prod_{i\geq 1} \left(e^{0} d\Lambda_i(t)\right)^{dN_i^D(t)} \left(1 - e^{0} dH_i(t)\right)^{1-dN_i^D(t)}} \\ &= \prod_{i\geq 1} \frac{\left(e^{\hat{\theta}_t}\right)^{dN_i^D(t)} \left(1 - e^{\hat{\theta}_t} d\Lambda_i(t)\right)^{1-dN_i^D(t)}}{\left(1 - d\Lambda_i(t)\right)^{1-dN_i^D(t)}} \end{aligned}$$

where $\hat{\theta}_t$ is the MLE over $\theta > 0$ at time t. We will discuss how to determine this estimate in 6.2.1. Once again using a repeated conditioning argument, the likelihood based on the information up to time t can be calculated using equation (6) and is equal to:

$$LR_{GLR}(t) = \prod_{i\geq 1} \frac{\left(e^{\hat{\theta}_t}\right)^{dN_i^D(t)} e^{-e^{\hat{\theta}_t}\Lambda_i(T_i)}}{e^{-\Lambda_i(T_i)}}.$$

Thus the log likelihood ratio up to time t is given by:

$$\ln(LR_{GLR}(t)) = \hat{\theta}_t N^D(t) - (e^{\hat{\theta}_t} - 1) \sum_{i\geq 1} \Lambda_i(t).$$

We can determine the ML estimate of θ at time t, which will be done in the following lemma.

Lemma 6.2.1. *The maximum likelihood estimate of θ for the Generalized likelihood ratio test introduced in section 6.2 is given by:*

$$\hat{\theta}_t = \max\left(0, \ln\left(\frac{N_D(t)}{\sum_{i\geq 1} \Lambda_i(t)}\right)\right). \tag{32}$$

Proof. To determine the maximum likelihood of θ, we first remind ourselves that we are determining the maximum likelihood estimator for θ for the hypotheses $\mu = 0$ against $\mu = \theta$ with θ unknown. Heuristically this means that we are testing whether the true *intensity* $e^{\mu} d\Lambda_i(t)$ differs from the national rate $d\Lambda_i(t)$, which was pre-determined from some sort of training set (a data set which we know to have in control procedures). Whenever we experience a failure (revision) we obtain some information as to how likely it was that the individual in question would experience a failure at that time point, this will then change the value of the MLE accordingly. Consider the likelihood up to time t using equation (7):

$$L(\theta|t) = \prod_{i\geq 1} \left(e^{\theta} d\Lambda_i(t)\right)^{dN_i^D(t)} e^{-e^{\theta}\Lambda_i(T_i)}$$

where we have n observations which are either active or have had an outcome or had their outcome censored. The logarithm of the likelihood (up to time t) is then given by:

$$l(\theta) = \sum_{i \geq 1} (dN_i^D(t))(\theta + \ln(d\Lambda_i(t))) - e^{\theta}\Lambda_i(t).$$

Taking the derivative w.r.t. θ and equating to zero we then obtain:

$$\sum_{i \geq 1} (dN_i^D(t)) - e^{\theta} \sum_{i \geq 1} \Lambda_i(t) = 0$$

which yields:

$$e^{\theta} = \frac{N^D(t)}{\sum_{i \geq 1} \Lambda_i(t)}.$$

The MLE for θ at time t is then given by:

$$\hat{\theta}_t = \max\left(0, \ln\left(\frac{N_D(t)}{\sum_{i \geq 1} \Lambda_i(t)}\right)\right)$$

where the cut-off at zero arises from the fact that we test the hypothesis of $\mu = 0$ against a hypothesis of $\mu = \theta > 0$. □

6.3 Definition

We summarise the results from the previous subsection by defining the Continuous Time Generalized Likelihood Ratio chart:

Definition 6.3.1. *The Continuous Time Generalized Likelihood Ratio chart (CTGLR) is given by:*

$$GLR(t) = \hat{\theta}_t N^D(t) - (e^{\hat{\theta}_t} - 1) \sum_{i \geq 1} \Lambda_i(t). \tag{33}$$

This chart is used to test the hypothesis that the cumulative intensity at an institution differs by a factor of e^{θ} (with $\theta > 0$ unknown) from the risk-adjusted null cumulative intensity $\Lambda_i(t)$, this can be stated as:

$$H_0 : \mu = 0$$
$$H_1 : \mu = \theta$$

The counting processed are as defined in section 5.1. The maximum likelihood estimator $\hat{\theta}_t$ was found in Lemma 6.2.1:

$$\hat{\theta}_t = \max\left(0, \ln\left(\frac{N^D(t)}{\sum_{i \geq 1} \Lambda_i(t)}\right)\right). \tag{34}$$

*The null hypothesis is rejected at time t when $GLR(t) \geq h$ for some $h > 0$, called the **control limit**.*

Unfortunately, the CTGLR requires more computational power than the CTCUSUM defined in Definition 5.3.1 as we have to determine the MLE of θ at every relevant timepoint, and re-determine the value of the statistic, instead of recursively determining the value of the chart. The main advantage

over the CTCUSUM chart is that the CTGLR chart no longer requires us to specify a value for the alternative hypothesis. Heuristically one could see the CTGLR as an automated CTCUSUM chart, but it is important to note that **they test different hypotheses**. A generalization of the CTGLR chart which tests the same hypothesis as the CTCUSUM chart will be stated in Definition 6.9.1.

Even though the CTGLR "automatically" determines a value of θ by means of an MLE, we still need to choose a value for the detection threshold h, which means that implicitly we will be choosing a value of μ at which we would like the chart to signal as rapidly as possible through this h. To let $\hat{\theta}_t$ converge towards the true value of μ, a sufficient amount of observations are needed, which means that the chart needs time to converge. In comparison, when the value of θ is not chosen in accordance with the true distribution of μ, the CTCUSUM chart may experience detection delays or high false alarm rates. Why this happens and the implications this has will be discussed in section 6.10.

6.3.1 Closed expression & non-negativity

Using definition 6.3.1, we can write a closed-form expression for $GLR(t)$. First of all, note that the MLE $\hat{\theta}_t$ will be equal to zero when $N_D(t) \leq \sum_{i=1}^{n} \Lambda_i(t)$. This happens when the contribution $\sum_{i=1}^{n} \Lambda_i(t)$ of the individuals who are active or have had a revision/censored observation is greater than the total amount of revisions up to that time point. Heuristically this happens when the institution considered has a lot of active cases and not many failures, or if we consider only one individual it happens when this individual has failed at or after time t for which $\Lambda_i(t) = 1$, which will later on be shown analytically in section 6.8. When $\hat{\theta}(t) = 0$ we clearly have $GLR(t) = 0$, thus if we would like for the chart to never rise above 0 the institution would need to have a sufficient ratio of new active (non-failing) patients compared to unfavourable outcomes combined with a sufficient amount of patients failing only when their cumulative hazard ratio has reached a sufficiently large number. To obtain a general expression we can substitute equation (34) into equation (33) to obtain:

$$GLR(t) = \begin{cases} N^D(t)\left(\ln\left(\frac{N^D(t)}{\sum_{i\geq 1}\Lambda_i(t)}\right) - 1\right) + \sum_{i=1}^{n}\Lambda_i(t), & \text{if } N^D(t) > \sum_{i\geq 1}\Lambda_i(t) \\ 0, & \text{if } N^D(t) \leq \sum_{i\geq 1}\Lambda_i(t). \end{cases} \tag{35}$$

Note that the chart is continuous in its argument, as the upper part of the equation above is also equal to zero when $N^D(t) = \sum_{i\geq 1}\Lambda_i(t)$. The chart will stay at zero when either patients fail at reasonable times or when failures do not happen proportionally faster than the influx of new patients. As $\hat{\theta}_t$ is determined over all time-points, a disadvantage is that the chart cannot quickly adjust to the current situation when the outcomes observed in the past were failing at or slower than expected under the baseline hazard. In this case the GLR chart can build up a "buffer" as $\sum_{i\geq 1}\Lambda_i(T_i)$ can become (very) large compared to $N^D(t)$. This is not the case for the CTCUSUM chart as it has a cut-off at zero and θ is constant. The CTCUSUM chart however can jump up only by the pre-defined value of θ, while the CTGLR chart can make way larger jumps, depending on the rate of failure in the (immediate) past. When there has been a period where a lot of unfavourable outcomes have been observed, both the CTCUSUM and CTGLR chart will need time to drift downwards. The CTCUSUM chart will drift downward depending on the chosen value of θ, while for the CTGLR chart the angle of downward drift depends on the current MLE of θ.

We show that the CTGLR chart is non-negative in the following lemma.

Lemma 6.3.1. *The CTGLR chart as defined in 6.3.1 is non-negative for every $t \geq 0$. The chart is strictly positive whenever $N^D(t) > \sum_{i\geq 1}\Lambda_i(t)$.*

Proof. When $N^D(t) \leq \sum_{i\geq 1}\Lambda_i(t)$ we have that $GLR(t) = 0 \geq 0$ from equation (35), thus the

statement holds. When $N^D(t) > \sum_{i\geq 1}\Lambda_i(t)$ we rewrite the upper part of equation (35) to obtain:

$$GLR(t) = N^D(t)\left(\ln\left(\frac{N^D(t)}{\sum_{i\geq 1}\Lambda_i(t)}\right) + \frac{\sum_{i\geq 1}\Lambda_i(t)}{N^D(t)} - 1\right).$$

Now rewriting the term between brackets using $A = N^D(t)$ and $B = \sum_{i\geq 1}\Lambda_i(T_i)$ we obtain:

$$\ln\left(\frac{A}{B}\right) + \frac{B}{A} - 1 = \frac{B}{A} - \left(1 + \ln\left(\frac{B}{A}\right)\right).$$

Now define $x := \ln\left(\frac{B}{A}\right)$. Because $A > B$ we thus obtain that $x > 0$. Using the Bernoulli inequality we obtain that:

$$1 + \ln\left(\frac{B}{A}\right) = 1 + x \leq e^x = \frac{B}{A}.$$

It can easily be seen that $1 + x = e^x$ if and only if $x = 0$. We obtained from our assumption that $x > 0$, therefore above inequality is strict. We conclude that $GLR(t) > 0$. □

Thus the CTGLR chart can only be non-negative, with the chart being equal to zero only when $N^D(t) \leq \sum_{i\geq 1}\Lambda_i(t)$. This is a consequence of the fact that we take the MLE as a maximum over 0 and the logarithm of the counting process divided by the total accumulated cumulative hazard. Therefore both the CTCUSUM chart and the CTGLR chart have a (sort of) cut-off at zero. This makes it much easier to compare the charts, but also makes it easy to misinterpret what it means when each chart is equal to zero. Whereas the CTCUSUM chart has a "hard" cut-off at zero, the CTGLR chart can have built up a big buffer in the value of the MLE $\hat{\theta}_t$, meaning that many failures are needed before the chart is able to rise above zero again.

6.4 Some proofs

In this section we build up some theory in order to prove our main result in section 6.5. We will build on results from the proof in Kalbfleisch & Biswas [2], which was presented in full in this thesis in section 5.4. First we will restate some of the results from this section, adjusting them on the go so they can be applied to the CTGLR. A few proofs will use results stated in the Appendix 15.

6.4.1 Expected value of the cumulative intensity

A term which we have often encountered in this thesis is $\sum_{i\geq 1}\Lambda_i(t)$. In this section we again name this term $A(t)$ and determine its expected value, similarly to section 5.4.

Lemma 6.4.1. *Assume that f_i^μ and h_i^μ are non-negative and Borel measurable. Define $A(t) = \sum_{i\geq 1}\Lambda_i(t)$, with $\Lambda_i(t)$ as in section 6.1. Then:*

$$\begin{aligned}\mathbb{E}[dA(u)] &= e^{-\mu}\psi\mathbb{E}_{Z_i}\left[F_i^\mu(u)\right]du \\ &=: \gamma_u du\end{aligned}$$

and:

$$\mathbb{E}[A(t)] = \int_0^t \gamma_u du$$

with ψ the rate of arrivals and $e^\mu\Lambda_i(t)$ the true risk-adjusted cumulative intensity of failure at the institution of interest.

Note that $A(t)$ is an infinite sum, but $\Lambda_i(t)$ is defined so that only "active" indices i are considered, see section 6.1.

Proof. Again we consider an institution with hazard rate e^μ times the baseline hazard rate. We have that:

$$U_{GLR}(t) := \hat{\theta}_t N^D(t) + (e^{\hat{\theta}_t} - 1)A(t) \tag{36}$$

where:

$$\hat{\theta}_t = \max\left\{0, \ln\left(\frac{N^D(t)}{\sum_{i\geq 1}\Lambda_i(t)}\right)\right\}.$$

Once again we assume that patients arrive according to a homogeneous Poisson process with rate $\psi > 0$. This time we will not study the process in equilibrium, instead choosing to take a more general approach by not limiting the qualifying outcomes to some period after the primary procedure, but instead choosing to consider their lifetime up until time of failure (or censoring). We repeat the calculation of the expected value of $dA(u)$ in section 5.4, only now slightly adjusted to the new situation:

$$\begin{aligned}\mathbb{E}[dA(u)] &= \mathbb{E}\left[\sum_{i\geq 1} Y_i(u)h_i(u-S_i)du\right] \\ &= e^{-\mu}\mathbb{E}\left[\sum_{i\geq 1} \mathbb{1}\{S_i \leq u, X_i \geq u - S_i\}e^{\mu}h_i(u-S_i)du\right].\end{aligned}$$

Here we use that $T_i = S_i + X_i$. Then using the law of total expectation twice (conditioning first on Z_i and then on S_i):

$$\begin{aligned}\mathbb{E}[dA(u)] &= e^{-\mu}\sum_{i\geq 1}\mathbb{E}_{S_i}\left[\mathbb{E}_{Z_i}\left[\mathbb{E}\left[\mathbb{1}\{S_i \leq u\}\mathbb{1}\{X_i \geq u - S_i|Z_i\}e^{\mu}h_i(u-S_i|Z_i)du\right]|S_i\right]\right] \\ &= e^{-\mu}\sum_{i\geq 1}\mathbb{E}_{S_i}\left[\mathbb{1}\{S_i \leq u\}\mathbb{E}_{Z_i}\left[\mathbb{E}\left[\mathbb{1}\{X_i \geq u - S_i|Z_i\}\right]e^{\mu}h_i(u-S_i|Z_i)du|S_i\right]\right] \\ &= e^{-\mu}\sum_{i\geq 1}\mathbb{E}_{S_i}\left[\mathbb{1}\{S_i \leq u\}\mathbb{E}_{Z_i}\left[\mathbb{P}\left(X_i \geq u - S_i|Z_i\right)h_i^{\mu}(u-S_i|Z_i)du|S_i\right]\right] \\ &= e^{-\mu}\sum_{i\geq 1}\mathbb{E}_{S_i}\left[\mathbb{1}\{S_i \leq u\}\mathbb{E}_{Z_i}\left[S_i^{\mu}\left(u - S_i|Z_i\right)h_i^{\mu}(u-S_i|Z_i)du|S_i\right]\right] \\ &= e^{-\mu}\sum_{i\geq 1}\mathbb{E}_{S_i}\left[\mathbb{1}\{S_i \leq u\}\mathbb{E}_{Z_i}\left[f_i^{\mu}(u - S_i|Z_i)|S_i\right]\right]du \\ &= e^{-\mu}\sum_{i\geq 1}\int_0^u \mathbb{E}_{Z_i}\left[f_i^{\mu}(u-x)\right]\psi\frac{e^{-\psi x}(\psi x)^{i-1}}{(i-1)!}dxdu \\ &= e^{-\mu}\psi\int_0^u \mathbb{E}_{Z_i}\left[f_i^{\mu}(u-x)\right]e^{-\psi x}\sum_{i\geq 1}\frac{(\psi x)^{i-1}}{(i-1)!}dxdu.\end{aligned}$$

We obtain using Lemma 15.2.1:

$$\mathbb{E}[dA(u)] = e^{-\mu}\psi \int_0^u \mathbb{E}_{Z_i}\left[f_i^\mu(u-x)\right] du \tag{37}$$

$$= e^{-\mu}\psi \mathbb{E}_{Z_i}\left[F_i^\mu(u)\right] du. \tag{38}$$

We define, similarly to above, the notation:

$$\gamma_u := e^{-\mu}\psi \mathbb{E}_{Z_i}[F_i^\mu(u)] > 0. \tag{39}$$

We assumed that f_i^μ and h_i^μ are non-negative and Borel measurable, therefore using Lemma 15.3.1 we obtain that $\mathbb{E}[A(t)] = \int_0^t \gamma_u du$. □

Note that in this case, γ_μ is **not** a constant as it depends on u, whereas γ as defined in equation (26) was a constant. We can relate the two as follows:

$$\gamma_{u=C} = \gamma$$

meaning γ_μ is just a generalized form of γ.

In practice we will see that γ_u gives sharper bounds on the ARL which we will determine for the CTGLR chart. Using γ_u in the proof of section 5.4 is not possible as a key point of that proof relied on the fact that γ was a constant.

6.4.2 Distribution of $N^D(t)$

Now consider the hospital as a queue, we then obtain the following result.

Lemma 6.4.2. *Consider a hospital where primary procedures arrive according to a Poisson process with rate ψ and the waiting time to failure for person i after the procedure has risk-adjusted distribution function F_i^μ. Then $N^D(t)$ is a Poisson process with rate $\frac{e^\mu}{t}\int_0^t \gamma_u du$.*

Proof. As patients arrive according to a Poisson process, the amount of failures after a primary procedure $N^D(t)$ can be seen as a departure process. Proposition 3.2.1 then tells us that $N^D(t)$ is a Poisson process with rate $\frac{e^\mu}{t}\int_0^t \gamma_u du$. □

In summary, we have:

$$U_{GLR}(t) = \hat{\theta}_t N^D(t) - (e^{\hat{\theta}_t} - 1)\int_0^t \gamma_u du + E_t \tag{40}$$

with $E_t = (e^{\hat{\theta}_t} - 1)(\int_0^t \gamma_u du - A(t))$, which is no longer a zero-mean process as $\hat{\theta}_t$ is not independent of $A(t)$. Because $\mathbb{E}[A(t)] = \int_0^t \gamma_u du$ (see Lemma 15.3.1) it is likely that E_t will stay small and therefore we approximate $E_t = 0$. Comparing equation (36) with equation (40) we notice that a consequence of this assumption is that we now assume that the only variability in the second part of $GLR(t)$ is in the $e^{\hat{\theta}_t}$ term, while ignoring the variability in $A(t)$. This will most likely lead to an underestimation in the variability of the chart, thus giving us unrealistically sharp bounds on the ARL.

Following the proof in section 5.4 we would want to use the property that $U_{GLR}(t)$ has stationary increments. However, $U_{GLR}(t)$ does not have stationary increments because $\hat{\theta}_t$ depends on t and we have that $\hat{\theta}_t - \hat{\theta}_s \neq \hat{\theta}_{t-s}$. Additionally, we no longer have information about the size of jumps the chart makes, as the chart no longer makes upward jumps of fixed size. Because of this we take another approach in the following sections.

6.4.3 Distribution of $\hat{\theta}_t$

The maximum likelihood estimate $\hat{\theta}_t$ is a key part of the CTGLR chart. We would therefore like to determine some properties of this estimate. To this end we use a well known property of maximum likelihood estimates to determine its asymptotic distribution. We are often more interested in the time t to detection, instead of the amount of patients n until detection. Because of this we relate the two to each other so that the following result is valid for $t \to \infty$ as well. In the following Lemma we use a result from Hoadley [27], which requires the model to adhere to some conditions. In practice, these conditions are mostly satisfied when f_i^μ is continuous and twice differentiable.

Lemma 6.4.3. *Let $t > 0$ and suppose we have n patients. Let $\{f_i^\mu, i = 1, ..., n\}$ meet conditions (N1)-(N9) of Hoadley [27]. Then, as $n \to \infty$:*

$$\sqrt{n}(\hat{\theta}_t - \mu) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{\overline{I(\mu, t)}}\right)$$

with $\overline{I(\mu,t)} = \frac{I(\mu,t)}{n}$ and $I(\mu, t)$ the Fisher information in all observations at time t. Moreover, assuming that $n = \psi \cdot t$ we have that:

$$\sqrt{t}\left(\hat{\theta}_t - \mu\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{\psi \cdot \overline{I(\mu, t)}}\right)$$

as $t \to \infty$.

Proof. Section 4 of Hoadley [27] tells us that under conditions $(N1) - (N9)$ as stated in the article our first statement holds. Most of these conditions are likely to hold when h_i^μ is continuous and twice differentiable. Finally, as patients arrive according to a Poisson process with rate ψ, it is reasonable (especially when ψ is large) to assume that $n = t \cdot \psi$, as we expect to see ψ patients arrive per time unit. Using this relation we then have that $n \to \infty$ implies that $t \to \infty$ (remember that ψ is constant), therefore using the properties of a normal distribution we obtain the second statement. □

6.4.3.1 Fisher information

In Lemma 6.4.3 we determined the asymptotic distribution of $\hat{\theta}_t$, which depends on the Fisher information. In the following lemma we determine the Fisher information for this model.

Lemma 6.4.4. *The Fisher information in all observations at time $t > 0$ is given by:*

$$I(\theta, t) = \psi \int_0^t \mathbb{E}_{Z_i}\left[F_i^\theta(k)\right] dk. \tag{41}$$

Proof. Similarly to Lemma 6.2.1 our likelihood function is given by:

$$L(\theta|t) = \prod_{i \geq 1} \left(e^\theta d\Lambda_i(t)\right)^{dN_i^D(t)} e^{-e^\theta \Lambda_i(t)}.$$

The log likelihood ratio is then given by:

$$l(\theta|t) = \sum_{i \geq 1} (dN_i^D(t))(\theta + \ln(d\Lambda_i(t))) - e^\theta \sum_{i \geq 1} \Lambda_i(t).$$

Taking the derivative w.r.t. θ yields:

$$\frac{\partial l(\theta|t)}{\partial \theta} = \sum_{i\geq 1} dN_i^D(t) - e^\theta \sum_{i\geq 1} \Lambda_i(t).$$

And the second derivative:

$$\frac{\partial^2 l(\theta|t)}{\partial \theta^2} = -e^\theta \sum_{i\geq 1} \Lambda_i(t).$$

Finally, the Fisher information is given by:

$$\begin{aligned} I(\theta,t) &= -\mathbb{E}\left[-e^\theta \sum_{i\geq 1} \Lambda_i(t) \mid \theta\right] && (42)\\ &= e^\theta \mathbb{E}_\theta\left[A(t)\right] && (43)\\ &= e^\theta e^{-\theta}\psi \int_0^t \mathbb{E}_{Z_i}\left[F_i^\theta(k)\right] dk && (44)\\ &= \psi \int_0^t \mathbb{E}_{Z_i}\left[F_i^\theta(k)\right] dk && (45) \end{aligned}$$

where we have used our result from Lemma 6.4.1. □

6.5 Main result: asymptotic distribution of the CTGLR

In this section we determine the asymptotic distribution of the continuous time GLR chart as defined in 6.3.1. This is the main theoretical result of this thesis, from which an approximation to the ARL of the CTGLR will follow in section 6.6 and consecutively an approximation to the ARL of the CTMAXGLR and CTCUSUM in section 6.9.

Theorem 6.5.1. *Consider an institution where patients arrive according to a Poisson process with rate ψ. Let h_i^μ be the risk-adjusted hazard rate of patient i and suppose the associated distribution functions $\{f_i^\mu, i = 1, ...\}$ meet conditions (N1)-(N9) of Hoadley [27] and let $\mu > 0$ be fixed. Suppose $A(t) = \sum_{i\geq 1} \Lambda_i(t)$ is constant with value $\mathbb{E}_\mu[A(t)]$ and that $N^D(t) = e^{\hat{\theta}_t} A(t)$. Then the GLR chart as defined in 6.3.1 converges in distribution:*

$$\sqrt{t}\left(GLR(t) - \left(\mu + e^{-\mu} - 1\right) I(\mu,t)\right) \xrightarrow{d} \mathcal{N}\left(0, t\mu^2 I(\mu,t)\right), \text{ when } \mu > 0 \quad (46)$$

$$t \cdot GLR(t) \xrightarrow{d} \frac{t\cdot\chi_1^2}{2} = \Gamma\left(\frac{1}{2}, t\right), \text{ when } \mu = 0 \quad (47)$$

as $t \to \infty$ where

$$I(\mu,t) := \psi \int_0^t \mathbb{E}_{Z_i}\left[F_i^\mu(k)\right] dk = e^\mu \int_0^t \gamma_u du \quad (48)$$

and χ^2 and Γ represent the Chi-squared and Gamma distributions respectively.

Note that assuming that $N^D(t) = e^{\hat{\theta}_t} A(t)$ when $\mu > 0$ is not very problematic, as we instantly recover this equality from equation (32) whenever $\hat{\theta}_t > 0$. As $\hat{\theta}_t$ converges towards the true value of μ, we are very likely to have $\hat{\theta}_t > 0$ whenever $\mu > 0$, especially for μ (and ψ) large.

Proof. **1. When $\mu > 0$:** Assuming that $N^D(t) = e^{\hat{\theta}_t} A(t)$ we can write:

$$\begin{aligned} GLR(t) &= \hat{\theta}_t N^D(t) - (e^{\hat{\theta}_t} - 1) \int_0^t \gamma_u du \\ &= \hat{\theta}_t e^{\hat{\theta}_t} A(t) - (e^{\hat{\theta}_t} - 1) \int_0^t \gamma_u du. \end{aligned}$$

Now assume that $A(t)$ is a constant with value $\mathbb{E}[A(t)] = \int_0^t \gamma_u du = I(\mu, t)$ (using Lemma 6.4.1). Then consider the function ϕ defined as:

$$\phi(x) = xe^x e^{-\mu} I(\mu, t) - e^x e^{-\mu} I(\mu, t) + e^{-\mu} I(\mu, t)$$

and note that $\phi(\hat{\theta}_t) = GLR(t)$. Additionally, we find that ϕ is differentiable at μ with derivative:

$$\phi'(x) = e^{-\mu} I(\mu, t) x e^x.$$

Then by the delta method (see section 7 of Van der Vaart. [28]) and the result of Lemma 6.4.3 we obtain that:

$$\sqrt{n} \left(\phi(\hat{\theta}_t) - \phi(\mu) \right) \xrightarrow{d} \mathcal{N} \left(0, \frac{(\phi'(\mu))^2}{I(\mu, t)} \right)$$

as $n \to \infty$ which reduces to:

$$\sqrt{t} \left(GLR(t) - \left(\mu + e^{-\mu} - 1 \right) K_t \right) \xrightarrow{d} \mathcal{N} \left(0, t\mu^2 I(\mu, t) \right)$$

when $t \to \infty$ using a similar argument ($n = \psi \cdot t$) as in the proof of Lemma 6.4.3.

2. When $\mu = 0$: In this case we can no longer use the delta method to determine the distribution of $GLR(t)$ as $\phi'(\mu) = \phi'(0) = 0$. Luckily we can use the second-order delta method (see Theorem 5.5.26 of Casella & Berger [29]). Note that:

$$\phi''(x) = (x + 1) e^x e^{-\mu} I(\mu, t)$$

and $\phi''(0) = I(\mu, t)$. Now the second-order delta-method in combination with Lemma 6.4.3 tells us that:

$$n \left(GLR(t) - \phi(0) \right) \xrightarrow{d} \frac{1}{I(\theta, t)} \frac{\phi''(0)}{2} \chi_1^2$$

as $n \to \infty$ which simplifies to:

$$t \cdot GLR(t) \xrightarrow{d} \frac{t}{2} \chi_1^2 = \Gamma \left(\frac{1}{2}, t \right)$$

as $t \to \infty$, using the scale parametrization of the Gamma distribution (and using that $n = \psi \cdot t$). □

Having determined an asymptotic distribution for the CTGLR, we can now approximate an (asymptotic) average run length, which will be done in section 6.6. Note that the result only holds for $t \to \infty$ or similarly $n \to \infty$, but we will see in section 7.1.2 that this result is quite applicable for small t as well, especially when ψ is large as $n = \psi \cdot t$.

6.5.1 Closed distribution for the CTGLR

It should be possible to obtain an even better analytic expression for the (asymptotic) distribution of the CTGLR, although it will most likely be quite a challenge. For this we note that we know the (asymptotic) distributions of almost all the separate terms of the CTGLR, as in section 6.4 we have shown the individual components of the CTGLR to have the following distributions:

$$GLR(t) = \underbrace{\hat{\theta}_t}_{\mathcal{N}(\mu, I(\theta,t)^{-1})} \cdot \underbrace{N^D(t)}_{\text{Poiss}\left(e^{\mu}\int_0^t \gamma_u du\right)} - \underbrace{\left(e^{\hat{\theta}_t} - 1\right)}_{\log\mathcal{N}(\mu-1, I(\theta,t)^{-1})} \int_0^t \gamma_u du$$

where we assumed that $A(t)$ is fixed with value $\int_0^t \gamma_u du$. It should be possible to determine the distribution of the whole expression. The biggest difficulty when doing so is determining the covariance of $\hat{\theta}_t$ and $N^D(t)$.

6.6 An approximation to the ARL

In this section we discuss how to determine an approximate ARL using the results above.

6.6.1 Using the main result

We discuss how to obtain an approximate ARL for the CTGLR chart using the main result of this thesis.

Corollary 6.6.1. *Consider an institution with risk-adjusted hazard rate h_i^{μ} and take $h > 0$. Then using the result from Theorem 6.5.1 one can obtain an approximate out of control average run length by solving the following equation for t:*

$$\mathbb{E}[GLR(t)] = (\mu + e^{-\mu} - 1)I(\mu, t) = h. \tag{49}$$

The obtained solution for t, denoted by $ARL_{GLR}(\mu, h)$, will then be an approximation of the average run length under the alternative hypothesis of the value $\mu > 0$ of interest.

Above equation can be solved either algebraically or numerically. Using a root finding algorithm like Newton-Raphson is convenient as we can use the fundamental theorem of calculus when the hazard is chosen parametrically.

6.6.2 Substituting confidence bounds

Instead of using the delta method to find an asymptotic distribution of the CTGLR chart we can use another approach to find confidence intervals for the CTGLR. Fix $t > 0$ large enough, then using Lemma 6.4.3 we have that:

$$\hat{\theta}_t \dot{\sim} \mathcal{N}\left(\mu, \frac{1}{I(\mu, t)}\right).$$

Additionally, in Lemma 6.4.2 we found that:

$$N^D(t) \sim \text{Poiss}\left(\psi \int_0^t \mathbb{E}\left[F_i^{\mu}(u)\right] du\right).$$

At every time-point $t > 0$ we can therefore find an (approximate) $1 - \alpha$ confidence level bounds for both $\hat{\theta}_t$ as well as for $N^D(t)$. We know that $\hat{\theta}_t$ and $N^D(t)$ have a strictly positive correlation, as

$\hat{\theta}_t = \ln\left(\frac{N^D(t)}{A(t)}\right)$, therefore we know that the lower confidence bound for $\hat{\theta}_t$ is associated with the lower bound for $N^D(t)$. If we now substitute these confidence bounds into:

$$GLR(t) = \hat{\theta}_t N^D(t) - (e^{\hat{\theta}_t} - 1)\int_0^t \gamma_u du$$

we can obtain approximate $1-\alpha$ confidence intervals for the CTGLR, even when $\mu = 0$.

6.7 Ignoring the covariates

We have seen that the term $\mathbb{E}_{Z_i}\left[F_i^{\mu}(k)\right]$ is present in many equations above. This is because we have a dependence on covariates, which we try to integrate out to obtain some theoretical expression. Often we are not interested in considering risk-adjustment terms, as we do not know what kind of distribution the risk-adjustment variables will have in our application. In the subsections below we will therefore sometimes ignore the risk-adjustment terms. When we do we simply replace $\mathbb{E}_{Z_i}\left[F_i^{\mu}(k)\right]$ with $F^{\mu}(k)$, as this is the equivalent when no risk-adjustment is present.

6.8 Asymptotic bounds

In this section we determine an asymptotic upper and lower bound for the value of the non risk-adjusted CTGLR chart.

Lemma 6.8.1. *Suppose an institution has had $n \in \mathbb{N}_{>0}$ primary procedures and the risk-adjusted hazard for failure is given by h_i^{θ}. If we only consider these patients, the expected value of the continuous time GLR chart can be bounded as follows when $t \to \infty$:*

$$n\left(\theta - (1-e^{-\theta})\right) \leq \mathbb{E}_{\theta}\left[\lim_{t\to\infty} GLR(t)\right] \leq -\sum_{i=1}^{n}\mathbb{E}_{\theta}\left[\ln(H_i(X_i))\right] - n + ne^{-\theta}. \tag{50}$$

To the best of our knowledge, further simplification of the expression $\mathbb{E}_{\theta}\left[\ln(H_i(X_i))\right]$ is not possible in general, but it should not be difficult to determine it computationally. Additionally, when we do not consider risk-adjustment terms, we have that $\sum_{i=1}^{n}\mathbb{E}_{\theta}\left[\ln(H_i(X_i))\right] = n\cdot\mathbb{E}_{\theta}\left[\ln(H_0(X_1))\right]$, which should not be too hard to determine analytically under some specific parametric assumptions.

Proof. First let us look at the expected value of the chart in the limit $t\to\infty$ having observed all n available outcomes, with n a positive integer which we fix beforehand. For this we want to determine the expected value of equation (35). The expectation of the lower part is trivial, for the upper part we would first like to determine the expectation of the sum of the cumulative hazards. Suppose that we have no censored observations, then $T \sim F_i(t) = 1 - S_i(t)$. Consider the hypotheses $\mu = 0$ against $\mu = \theta$, then using integration by parts we obtain:

$$\begin{aligned}
\mathbb{E}_{\theta}[\lim_{t\to\infty}\Lambda_i(t)] = \mathbb{E}_{\theta}[H_i(X_i)] &= \int_0^{\infty} H_i(t) f_i^{\theta}(t)dt = \int_0^{\infty}\int_0^{t} e^{Z_i\beta}h_0(t)dt\cdot -\frac{\partial S_i^{\theta}(t)}{\partial t}dt && (51)\\
&= -\left(\left[H_i(t)S_i^{\theta}(t)\right]_0^{\infty} - \int_0^{\infty} S_i^{\theta}(t)h_i(t)dt\right) && (52)\\
&= -\left(0 - 0 - e^{-\theta}\int_0^{\infty} S_i^{\theta}(t)h_i^{\theta}(t)dt\right) && (53)\\
&= e^{-\theta}\int_0^{\infty} f_i^{\theta}(t)dt = e^{-\theta} && (54)
\end{aligned}$$

which supports the heuristic statement made in the beginning of section 6.3.1 that the expected value of the cumulative hazard is equal to 1 for a person failing according to the baseline hazard rate. As we have a fixed number n of observations and no censoring we obtain:

$$\mathbb{E}_\theta[\lim_{t\to\infty} N^D(t)] = \mathbb{E}_\theta\left[\lim_{t\to\infty}\sum_{i=1}^{n} N_i^D(t)\right] = n.$$

Then using Jensen's inequality:

$$\begin{aligned}
\mathbb{E}_\theta\left[\lim_{t\to\infty}\ln\left(\frac{N^D(t)}{\sum_{i=1}^n H_i(T_i)}\right)\right] &= \ln(n) - \mathbb{E}_\theta\left[\ln\left(\sum_{i=1}^{n} H_i(X_i)\right)\right] \\
&\geq \ln(n) - \ln\left(\sum_{i=1}^{n}\mathbb{E}_\theta[H_i(X_i)]\right) \\
&= \ln(n) - \ln(ne^{-\theta}) \\
&= \theta.
\end{aligned}$$

Substituting the above results into (35) and assuming we have a total of n outcomes, all under H_θ, we expect the chart to converge towards:

$$\mathbb{E}_\theta\left[\lim_{t\to\infty} GLR(t)\right] \geq n\cdot\theta - n + ne^{-\theta} = n\left(\theta - (1-e^{-\theta})\right). \tag{55}$$

Conversely, we can find an upper bound by using the logarithm of the inequality of arithmetic and geometric means:

$$\ln\left(\frac{\sum_{i=1}^n x_i}{n}\right) \geq \frac{\sum_{i=1}^n \ln(x_i)}{n}$$

which yields:

$$\begin{aligned}
\mathbb{E}_\theta\left[\lim_{t\to\infty}\ln\left(\frac{N^D(t)}{\sum_{i=1}^n H_i(T_i)}\right)\right] &= -\mathbb{E}_\theta\left[\ln\left(\frac{\sum_{i=1}^n H_i(X_i)}{n}\right)\right] \\
&\leq -\mathbb{E}_\theta\left[\frac{\sum_{i=1}^n \ln(H_i(X_i))}{n}\right] \\
&= -\frac{\sum_{i=1}^n \mathbb{E}_\theta\left[\ln(H_i(X_i))\right]}{n}
\end{aligned}$$

where the expectation does not have a closed general form. Thus giving the following upper bound for the GLR:

$$\mathbb{E}_\theta\left[\lim_{t\to\infty} GLR(t)\right] \leq -\sum_{i=1}^{n}\mathbb{E}_\theta\left[\ln(H_i(X_i))\right] - n + ne^{-\theta}. \tag{56}$$

Because we consider the risk-adjusted case, the X_i are not i.i.d. and thus the summation also does not disappear. Note that it should still be possible to determine this upper bound numerically or calculate the value for some chosen (cumulative) hazard function. Combining both above results we thus have that:

$$n\left(\theta - (1-e^{-\theta})\right) \leq \mathbb{E}_\theta\left[\lim_{t\to\infty} GLR(t)\right] \leq -\sum_{i=1}^{n}\mathbb{E}_\theta\left[\ln(H_i(X_i))\right] - n + ne^{-\theta}.$$

□

6.9 Continuous time MAXGLR chart

In this section we consider the Continuous Time MAXGLR chart, which is a continuous time generalization of the discrete time maxGLR chart introduced in equation (15), similarly to how the CTGLR chart is a generalization of the discrete time GLR chart. The CTMAXGLR can be seen as the "automated" CTCUSUM chart, as it is used to test the same hypotheses, and determines a suitable value for θ using a ML estimate.

Entirely analogous to how the discrete time maxGLR chart was derived from the discrete time GLR chart in section 4.3, the CTMAXGLR chart can be derived by maximizing the CTGLR chart over the last k observations.

Definition 6.9.1. *The Continuous Time Maximized Generalized Likelihood Ratio chart (CTMAXGLR) is given by:*

$$MAXGLR(t) = \max_k \hat{\theta}_{k,t} N_k^D(t) - (e^{\hat{\theta}_{k,t}} - 1) \sum_{i=k}^{n} \Lambda_i(t) \tag{57}$$

where the subscript k indicates that the quantities are taken over the last $n-k$ patients. This chart is used to test the hypothesis that the cumulative intensity at an institution differs by a factor of e^θ (with $\theta > 0$ unknown) from the null cumulative intensity $\Lambda_i(t)$, starting from some unknown patient ν:

$$\begin{aligned} H_0 :& X_1, X_2, ... \sim \Lambda_i \\ H_1 :& X_1, ..., X_{\nu-1} \sim \Lambda_i \\ & X_\nu, X_{\nu+1}, ... \sim \Lambda_i^\theta \end{aligned}$$

The counting processed are as defined in section 5.1. The maximum likelihood estimator $\hat{\theta}_k$ (limited to the last $n-k$ patients) was found in Lemma 6.2.1:

$$\hat{\theta}_{k,t} = \max\left(0, \ln\left(\frac{N_k^D(t)}{\sum_{i\geq k} \Lambda_i(t)}\right)\right). \tag{58}$$

*The null hypothesis is rejected at time t when $MAXGLR(t) \geq h$ for some $h > 0$, called the **control limit**.*

Note that per definition the value of the CTMAXGLR is always bigger or equal to the value of the CTGLR chart. The maximum likelihood estimate $\hat{\theta}_t$ found in Lemma 6.2.1 can build up a buffer when failures happen according or slower than the null hazard rate, as $\sum_{i\geq 1} \Lambda_i(t)$ can become (much) larger than $N^D(t)$. This means that if at the beginning of observations failures happen slower or according to the null rate $\Lambda_i(t)$ the CTGLR chart will need some time to adjust after failures start happening faster than the null rate, which will lead to detection delays. To counteract this fact we can consider the CTMAXGLR, which does not have this problem as we eliminate the possibility of building up a buffer by maximizing over the last $n-k$ observations. This in turn makes the CTMAXGLR prone to false signalling when a few consecutive failures happen very quickly. Another problem is that whereas in a discrete time maxGLR the ordering of the outcomes is very clear, the ordering in a continuous time chart is not quite as evident. The most logical decision is to order the individuals by their time of entry into the study, therefore we order by S_i for every i. Then we have to make a decision on how to order for ties in the data, thus when multiple individuals enter the study at the same time point. In this case, we can choose to order them by their observation time (X_i) if it is known and either randomly or (when using risk-adjustment) by their risk factor when their outcome time is not known. The choice made at this point directly influences how conservative the chart will be. Leaving

out high-risk patients which have not experienced a failure for a relatively long time will increase the value of the CTMAXGLR whereas leaving out patients with low risk which have had a very early failure will decrease the value of the chart in that iteration. In any case, the CTMAXGLR will always be greater than the CTGLR, but the choice of ordering can have a significant impact on the in and out of control run length in the presence of many ties. If false alarms are not a big problem, then the most conservative ordering of the data should be chosen, as this leads to smaller detection delays.

Proposition 6.9.1. *Consider all processes which are in control for the first $\nu - 1$ observations, and out of control with excess hazard rate $e^{\mu} > 1$ afterwards. Denote the time of the ν-th observation by τ. Denote the average run length of the CTGLR chart with control limit h starting from time τ as $ARL_{GLR}(\mu, h, \tau)$. Denote the average run length of the continuous time MAXGLR chart from $t = 0$ as $ARL_{MAXGLR}(\mu, h)$. Then:*

$$ARL_{MAXGLR}(\mu, h) \leq ARL_{GLR}(\mu, h, \tau) + \tau.$$

Proof. At every time point, the CTMAXGLR chart is simply a CTGLR chart which ignores the first k observations, where k is chosen such that the CTMAXGLR obtains its maximum possible value. Per definition, this implies that at any time point the CTMAXGLR will always be greater or equal to the CTGLR chart, irrespective of the starting point of the CTGLR. This means that for the same control limit $h > 0$, the CTMAXGLR will always hit the control limit earlier or at the same time as a CTGLR chart starting at any time point. □

Heuristically this proposition can be interpreted as follows: Suppose a hospital is not initially out of control, the run length of the CTMAXGLR chart is then at most the run length of the CTGLR chart for that hospital considered from the time observations started going out of control, plus the time until observations start going out of control. This will be of great importance in section 7, where, using a simulation study, we will compare the run lengths of the CTCUSUM and CTGLR charts when the process is initially out of control. Using this proposition we can then obtain an upper boundary for the run length of the CTMAXGLR. This run length will then be meaningful in practical applications, where the process is not always initially out of control.

6.10 Theoretical comparison of ARL between CTCUSUM and CT(MAX)GLR

In this section we will compare the approximation of the ARL of the CTCUSUM chart obtained by Biswas & Kalbfleisch [2] in equation (30) with the approximation of the ARL which we found for the CTGLR chart in Corollary 6.6.1.

To derive the ARL of the CTCUSUM chart as stated in equation (30) Biswas & Kalbfleisch assumed that observations begin at time point $t = -C$ years, so that at time $t = 0$ process is in equilibrium. Unfortunately this assumption is quite problematic when C grows large, as in real life applications we cannot ignore the first C years of observations or assume that the process immediately starts in equilibrium. Because of this, their approximation for the ARL is overly optimistic whenever C grows large. We found that for $C = 1$, the approximation works quite well, even when considering outcomes from $t = 0$. For C larger than 1 however, we found that the approximate ARL was way too small compared to simulation results (see Table 3). Besides this, it is difficult to compare their theoretical ARL with the one we obtain from Corollary 6.6.1, as we consider outcomes not limited to C years post transplant. This means that in theory, our result in Corollary 6.6.1 should be comparable with the ARL in equation 30 when $C \to \infty$. We argued above however that when C becomes large, equation (30) is no longer valid therefore we cannot compare our theoretical results meaningfully in this way.

To still compare the average run lengths of the two charts, we can derive a result similar to Corollary 6.6.1 for the CTCUSUM chart when $C = \infty$, using the exact same steps as in sections 6.5 and 6.6.

Corollary 6.10.1. *Consider an institution with risk-adjusted hazard rate h_i^μ and choose $h > 0$. Then one can obtain an approximate out of control average run length for the CTCUSUM with $\theta > 0$ and $C = \infty$ such that $\theta + e^{-\mu} - \frac{e^\theta}{e^\mu} > 0$ by solving the following equation for t:*

$$\mathbb{E}[G_\theta(t)] = (\theta + e^{-\mu} - \frac{e^\theta}{e^\mu})I(\mu, t) = h. \tag{59}$$

The obtained solution for t, denoted by $ARL_{CUS}(\theta, \mu, h)$, will then be an approximation of the average run length of the CTCUSUM under the alternative hypothesis $\mu > 0$ for the chosen value of θ.

Proof. Taking the expected value of equation (23) while assuming that $A(t)$ is constant with value $\mathbb{E}[A(t)] = e^{-\mu}I(\mu, t)$ (see 6.4.1) we obtain the approximate expected value of the CTCUSUM chart. We recover an approximate ARL by equating the expected value to a pre-defined cut-off value $h > 0$. □

Note that $\mathbb{E}[G_\theta(t)]$ can become negative when $\theta > \mu$, this means that it is not always possible to solve for the ARL in this way. The result allows us to compare the expected values of the CTCUSUM and CTGLR charts respectively., which restating the results above are:

$$\mathbb{E}[G_\theta(t)] = (\theta + e^{-\mu} - \frac{e^\theta}{e^\mu})I(\mu, t) \tag{60}$$

$$\mathbb{E}[GLR(t)] = (\mu + e^{-\mu} - 1)I(\mu, t) \tag{61}$$

allowing us to make the following statement.

Corollary 6.10.2. *Consider an institution with cumulative intensity $e^\mu\Lambda_i(t)$ for some $\mu > 0$ and choose $h > 0$. The approximate average run length of the CTGLR chart as defined in 6.6.1 is smaller or equal than the approximate average run length of the CTCUSUM chart with value θ as defined in 6.10.1 for testing against a null cumulative intensity of $\Lambda_i(t)$ using the value of h chosen above. The average run lengths are equal when $\theta = \mu$.*

Proof. Fixing $h > 0$ we easily see from Corollaries 6.10.1 and 6.6.1 that the average run length of both charts depend solely on the expected value of the charts as stated in equation (60). If we can thus show that $\mathbb{E}[GLR(t)] \geq \mathbb{E}[G_\theta(t)]$ for any $t > 0$ with equality only when $\theta = \mu$, the statement will hold. Note that as $I(\mu, t)$ only depends on μ and t, we only have to show that $\theta + e^{-\mu} - \frac{e^\theta}{e^\mu} \leq \mu + e^{-\mu} - 1$ or alternatively:

$$\theta - \mu \leq e^{\theta-\mu} - 1.$$

This follows directly from Bernoulli's inequality ($1 + x \leq e^x$ for any $x \in \mathbb{R}$), with equality only when $\theta - \mu = 0$, or $\theta = \mu$. □

We therefore find that the CTGLR chart always has a shorter approximate average run length when considering hospitals which are initially out of control, and using Proposition 6.9.1 we find that the CTMAXGLR has an even shorter run length. Heuristically this can be seen as the CT(MAX)GLR automatically selecting the true value of μ for us, by using a maximum likelihood estimate. The ML estimate needs some time (or equivalently a sufficient number of observations) to converge towards the true value, therefore a CTCUSUM chart with the right value of θ will in reality have a shorter detection time if the process is out of control from the start. This also tells us that if the value of θ in the CTCUSUM chart is chosen not in accordance with the true value of μ, the chart will not have an optimal stopping time, making the CT(MAX)GLR chart preferable when we do not know the true value of μ or when the rate varies over time. Unfortunately this statement does not imply that the CT(MAX)GLR is better than the CTCUSUM chart even in this aspect, as the CT(MAX)GLR

chart can still have a shorter in control average run length, which is not desirable. As we do not have a reliable theoretical result for the in control ARL, we cannot make a definite statement about which chart is better when considering type I errors. Besides this, the results for the ARL are only approximate, and only valid for large values of n or t. This means that we do not expect our theoretical results to sufficiently describe situations in which detections happen at early time points (i.e. low value of the control limit h, very high rate of failure etc.).

7 A simulation study

In this section we perform a simulation study in which we compare (mainly on the average run length) the CTCUSUM chart as defined in 5.3.1 with the CT(MAX)GLR charts defined in 6.3.1 and 6.9.1. The CTCUSUM will be considered with $C = \infty$, therefore considering all revisions post primary procedure as qualifying. The context of the simulations will be purely medical, where we consider N hospitals all performing primary hip replacement surgery (called the primary procedure) on patients, at which point the patients enter into the study. Patients arrive according to a Poisson process (see 3.1) with rate ψ. The outcome of interest is revision of the prosthesis. We assume that there is some known (null) cumulative intensity $\Lambda_i(t)$ which indicates an average rate of failure. Hospitals then deviate from this intensity by a factor of e^{μ}, where $\mu > 0$ ($e^{\mu} > 1$) means that their quality is worse than average and $\mu \leq 0$ ($e^{\mu} \leq 1$) indicates that it is up to standard or better than average. We will not consider censoring or competing risk mechanisms in the simulation studies of this section. This is because we are trying to determine whether the asymptotic theoretical results (such as the ARL) from section 6 are also valid at relatively small time points for survival outcomes. By implementing censoring and competing risks mechanisms we would no longer be able to compare with said theoretical results, as the assumptions in section 6 would no longer hold causing the run lengths to be skewed. In a practical scenario, these mechanisms could be of great importance. For this reason, right-censoring mechanisms will be considered in section 13. Competing risks will not be considered at all, because expert (medical) knowledge is required on whether the inclusion of death (in this instance) into the chart is desirable at all.

In every section we will clearly indicate under which assumptions the simulations were performed, this will be done by specifying all quantities of interest. When no quantities are specified, the following table will be guiding, see Table 2.

Number of charts/hospitals	Distribution of time to failure	Deviation factor under H_1	Poisson arrival rate
N	Exponential	e^{μ}	ψ
3000	Rate $\lambda = 0.002$	1.4	2.28

Table 2: Standard simulation parameters employed in this section.

The **simulation procedure** employed in this and further sections is summarized in section 15.4. We choose $N = 3000$ so that we can expect enough precision in the results under a realistic restriction on the computation time. The rate was chosen so that about 50 percent of patients will have failed after the first year under the null hypothesis. This was also done out of computational considerations, as longer failure times require us to construct the charts for longer periods of time. In Figure 3 the survival curve and cumulative hazard rate of the exponential distribution with rate $\lambda = 0.002$ is shown. Around $t = 3000$ days (about 8 years) most of the patients will have failed, and the cumulative hazard rate is a linear function of time. The parameter μ was chosen such that the failure rate with respect to the null distribution was sufficiently large to be detectable, while not making the failures happen too fast. The rate ψ was chosen to be approximately equal to the rate at the largest hospital present in our data set (see section 12), so as to make the amount of patients comparable to real life applications.

7.1 Assessing the theoretical assumptions and results

In sections 6.4.3 and 6.5 we have used quite a lot of approximations. In this section we assess whether these approximations are realistic by means of simulation, using the values specified in Table 2.

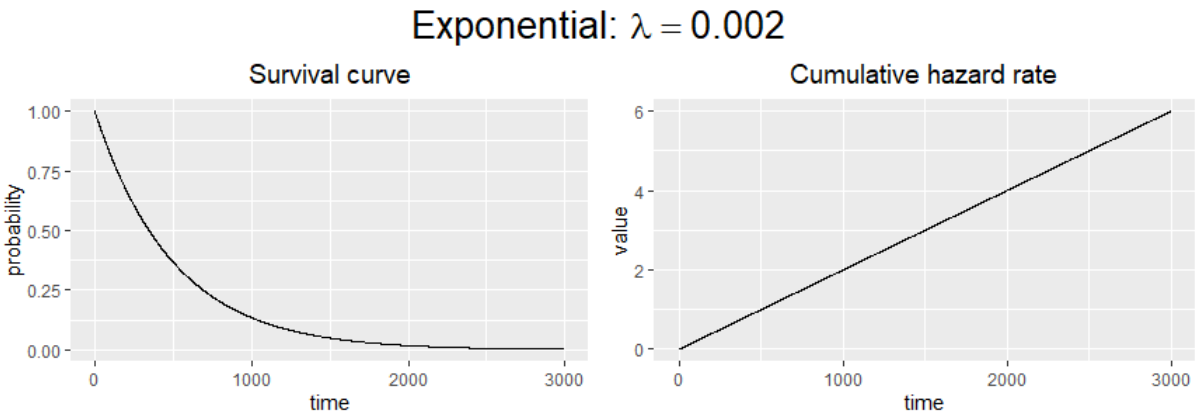


Figure 3: Survival curve (left) and cumulative hazard rate (right) for the exponential distribution with rate 0.002.

7.1.1 Asymptotic convergence of ML estimate

For the asymptotic convergence of the ML estimate in Lemma 6.4.2 we assumed that conditions $(N1) - (N9)$ from [27] were satisfied. All of these conditions are likely satisfied when the distribution function is continuous and twice differentiable.

We generate $N = 3000$ simulated hospitals according to Table 2 (see also section 15.4 step 3.) and evaluate the CTGLR on them. We plot the determined values of $\hat{\theta}_t$ for these CTGLR charts. Then we overplot the theoretical approximate 95 percent confidence intervals at every time point (fixing t and assuming it is large enough at every time-point) as well as the Monte Carlo 95 percent confidence intervals so that we can compare them. The result can be seen in Figure 4. This figure seems to reinforce our theoretical result, as the theoretical and Monte Carlo bounds coincide as t grows larger. Notable is how the monte carlo intervals are inside the theoretical intervals, indicating that convergence happens more rapidly than theoretically expected (in this case).

7.1.2 Confidence interval bounds for the CTGLR

We would like to asses whether our asymptotic distribution for the CTGLR chart found in Theorem 6.5.1 coincides with reality when t is not very large. Using the same $N = 3000$ CTGLR charts as above we construct approximate 95 percent confidence intervals for the value of these charts using both our main result 6.5.1 (fixing t and assuming t is large enough) and the result in section 6.6.2. In our main result we used the (second order) delta method to find an approximate distribution, while in section 6.6.2 we used a simple substitution argument. The resulting figures can be found in Figure 5. We can see that the bound substitution method gives slightly better bounds for early time points, whereas the delta method seems to underestimate both the upper as well as the lower MC confidence intervals for these early time points. This is because when programming the delta method we only cut-off the confidence intervals at zero after they have been generated from the normal distribution, whereas in bound substitution we use a cut-off for both $N^D(t)$ as well as for $\hat{\theta}_t$ at zero, before substituting them into the final expression. This is because both these values cannot (per assumption) be less than zero. We can see that both these theoretical bounds start overestimating the variance in the chart after a while, which is not in line with what we expected in section 6.4.2, as we do not consider the variability of $A(t)$ in this approximation. A possible reason for this is that in Theorem 6.5.1 we used the equality $N^D(t) = e^{\hat{\theta}_t} A(t)$ so that we could use the delta method. Afterwards we assumed $A(t)$ was a constant equal to its expected value. Most likely $e^{\hat{\theta}_t}$ has a bigger variance than $N^D(t)$, therefore yielding wider

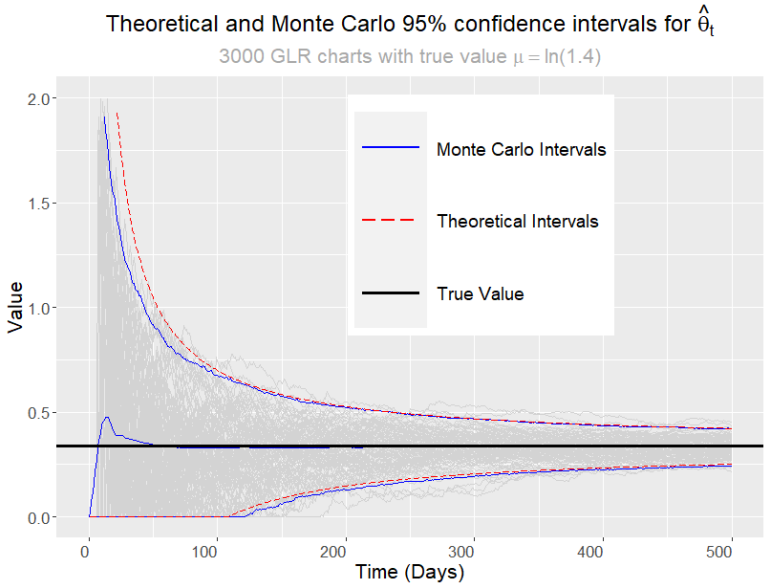
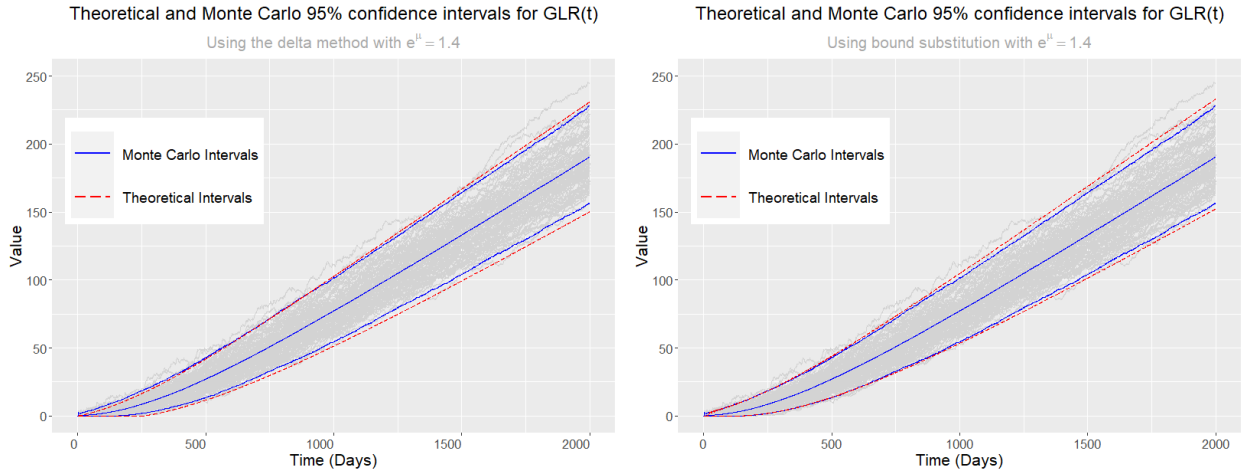


Figure 4: Theoretical and Monte Carlo 95 percent confidence intervals for $\hat{\theta}_t$ (Lemma 6.4.3). The figure was constructed using the parameters in Table 2. Whereas for the theoretical intervals only the upper and lower bounds are plotted, the MC mean is also included in the figures, as well as the true value of μ.

confidence intervals.

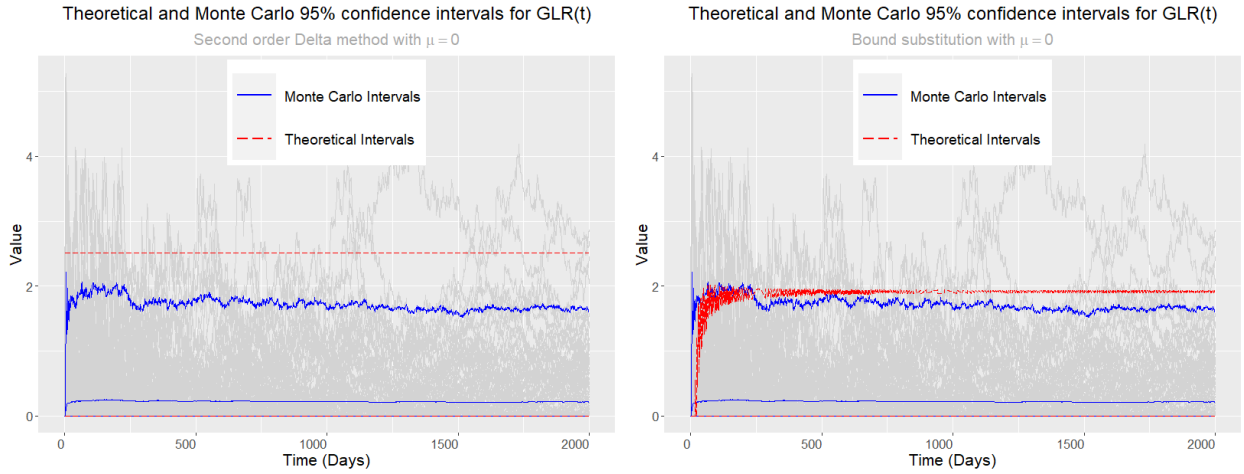


(a) Delta Method (b) Bound Substitution

Figure 5: Theoretical and Monte Carlo 95 percent confidence intervals for $U_{GLR}(t)$ using the (a) Delta Method and (b) bound substitution. Both figures were constructed using the parameters described in Table 2. Whereas for the theoretical intervals only the upper and lower bounds are plotted, the MC mean is also included in the figures.

Above we used a value of $e^{\mu} = 1.4$. We saw in Theorem 6.5.1 that when $e^{\mu} = 1$ the asymptotic distribution is different from when $e^{\mu} > 1$. The result from section 6.6.2 is unchanged however for $e^{\mu} = 1$. We therefore compare these two results again when $e^{\mu} = 1$. For this we use the same strategy as above to create Figure 6. We can see that using the second order delta method (main result) we

do not get sharp upper bounds compared to the bound substitution method in section 6.6.2. Besides this the upper bound using substitution seems to be slightly unstable, but not problematically so.



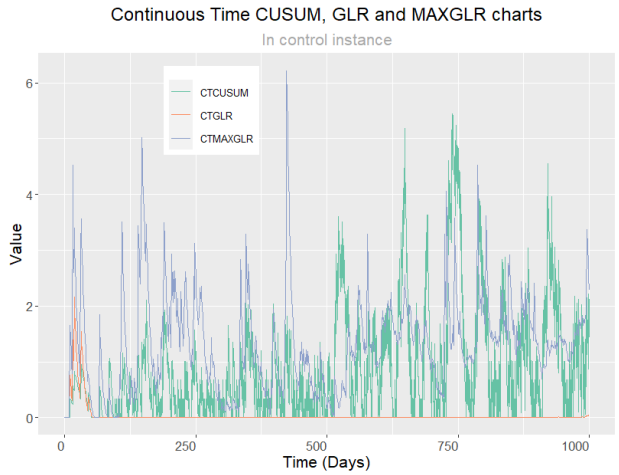
(a) Second Order Delta Method (b) Bound Substitution

Figure 6: Theoretical and Monte Carlo 95 percent confidence intervals for $U_{GLR}(t)$ using the (a) Second Order Delta Method and (b) bound substitution. Both figures were constructed using the parameters described in Table 2, but instead of $e^{\mu} = 1.4$ we now have $e^{\mu} = 1$. Whereas for the theoretical intervals only the upper and lower bounds are plotted, the MC mean is also included in the figures.

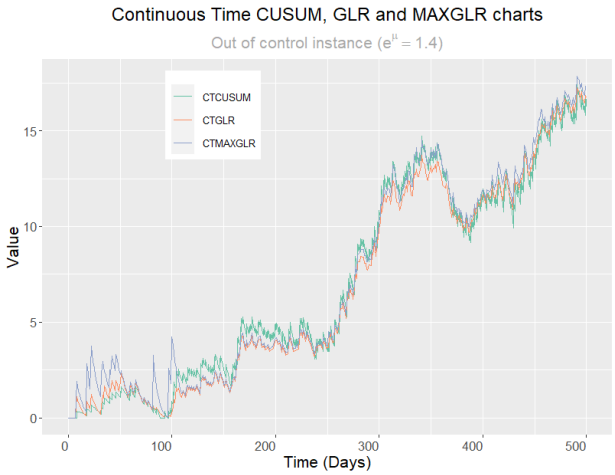
7.2 A short example

First of all, we would like to emphasize using a short simulation example that the CTGLR chart as defined in 6.3.1 is not suitable for testing for a delayed change in cumulative intensity, therefore stressing that it should only be compared with the CTCUSUM as defined in 5.3.1 when we know that the process is out of control from the start. When this is not the case, we should always compare the CTCUSUM with the CTMAXGLR as defined in 6.9.1. To show this we consider 3 simulated institutions. One where the process is in control during the whole observation time, one where the process is out of control during the whole observation time and one where the process is in control in the first 500 time points, and out of control starting from $t = 500$. We construct the CTCUSUM, CTGLR and CTMAXGLR for all 3 institutions, to obtain Figure 7.

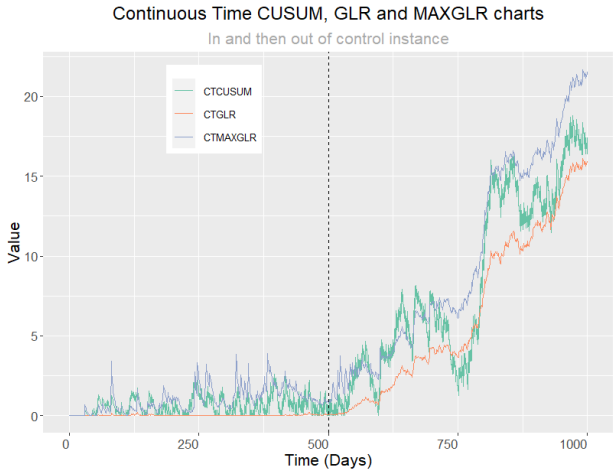
We can see in Figure 7a that the GLR stays at zero for most of the time, while the MAXGLR and CUSUM charts have values in a comparable range. This is because the CTGLR can build up a buffer in the value of $\hat{\theta}_t$ when the observations are in control, which is not the case for the CTCUSUM and CTMAXGLR. In Figure 7b we can see that all 3 charts take almost the same values at all time points. **This means that when the process is out of control the whole time, we can compare the CTCUSUM chart with the CTGLR chart**, instead of determining the CTMAXGLR chart, thus saving us a lot of computational power. Finally, in Figure 7c we can see that the CTMAXGLR and CTCUSUM take comparable values, while the CTGLR chart is lagging behind after the observations go out of control. This is once again because of the buffer the CTGLR can build up. Notable is how the CTCUSUM chart drops rapidly at the 600 day mark, while the CT(MAX)GLR experiences a period of stagnation instead of dropping. Afterwards we can see that the CTCUSUM chart catches up to the CTMAXGLR chart. Overall the CTCUSUM chart seems to be less stable than the CTGLR charts in these three examples.



(a) In control



(b) Out of control



(c) In and then out of control

Figure 7: One continuous time CUSUM, GLR and MAXGLR chart constructed for an (a) in control (b) out of control (c) in and then out of control institution. The out of control rate was chosen as $e^{\mu} = 1.4$, while the in control rate was $e^{\mu} = 1$. The dotted line in (c) indicates when the observations start going out of control.

7.3 Comparing the ARLs

As we saw in section 7.2 we can compare the CTCUSUM with the CTGLR when the observations are out of control from the beginning of the study. This allows us to perform quite extensive simulation studies, as both the CTCUSUM and CTGLR charts do not require a lot of computational power. We take the following approach in this section. Consider 2 CTCUSUM procedures, one with $e^{\theta} = 1.4$ and one with $e^{\theta} = 1.8$ and consider the CTGLR procedure. We generate 3000 institutions using Table 2, but with $e^{\mu} = 1$, thus under the null rate. We determine cut-off values h for all 3 charts such that their average run length under the null rate is 15 years, therefore we expect the charts to give a false signal every 15 years using the determined value of h. A summary of the simulation procedure is given in section 15.4.

We saw that the CTGLR chart can build up a buffer under the null rate, therefore we determined the value of h using the CTMAXGLR chart. As the CTMAXGLR takes very long to construct, we

determine the control limit on a small sample of $N = 120$ institutions, therefore introducing more variability into our results. Then once again we generate $N = 3000$ hospitals according to Table 2, with different values for μ and evaluate all three charts on them. We add the standard deviation of the run lengths, as well as the Median Run Length and the Theoretical value of the ARL as determined in Corollary 6.6.1 and Corollary 6.10.1. We add the theoretical value of the ARL with $C = 1$ year as determined by Biswas & Kalbfleisch [2], which can also be found in Equation (30). The results can be found in Table 3.

	CTCUSUM $e^\theta = 1.4, h = 6.82$			CTCUSUM $e^\theta = 1.8, h = 8.35$			CTGLR $h = 7.73$		
e^μ	ARL(SD)	MRL	Theor(B&K)	ARL(SD)	MRL	Theor(B&K)	ARL(SD)	MRL	Theor
1	5510 (4930)	4056	∞ (12086)	5478 (4739)	4104	∞ (16815)	5528 (4666)	4398	∞
1.2	409 (184)	374	1352 (274)	639 (366)	572	Inf (653)	480 (163)	474	511
1.4	205 (57)	198	227 (78)	240 (100)	223	490 (109)	229 (72)	228	243
1.6	152 (33)	148	159 (46)	153 (48)	145	177 (48)	153 (48)	151	162
1.8	127 (24)	125	130 (34)	119 (31)	116	128 (31)	117 (37)	117	123
2	110 (20)	109	112 (27)	101 (23)	99	106 (23)	95 (30)	94	100
2.2	99 (16)	98	101 (23)	89 (19)	87	92 (19)	81 (25)	80	85
2.4	91 (15)	91	92 (21)	81 (16)	80	82 (16)	71 (23)	71	74
2.6	85 (13)	84	85 (19)	74 (14)	73	75 (15)	63 (20)	62	65
2.8	79 (12)	79	80 (17)	69 (13)	68	70 (13)	57 (18)	57	59
3	75 (11)	75	75 (16)	65 (12)	64	66 (12)	52 (17)	51	54

Table 3: A comparison of ARLs (in days) between 2 CTCUSUM procedures ($e^\theta = 1.4, 1.8$) and the CTGLR, all with a value of h chosen such that the average run length under the null hypothesis is approximately 15 years. Evaluated on a sample of $N = 3000$ hospitals with true rate μ. Additionally, the standard deviation and Median Run Length are displayed, as well as the theoretical values for the ARL as determined in Corollary 6.6.1 and Corollary 6.10.1. Finally, we also display the theoretical ARL as determined by Biswas & Kalbfleisch [2] in equation (30) with $C = 1$ year, denoted by (B&K).

A few interesting things can be seen in the table, which are summed up below.

- Notice how the value of h for the CT(MAX)GLR lies between the values of h for the CTCUSUM procedures. Unfortunately it was not possible to take the in control ARL to be exactly the same for all 3 charts, due to computational limitations. Notice how the CTGLR chart has the largest in control run length, meaning that out of the 3 charts, the CTGLR chart has the most conservative value for h.

- The CTCUSUM with $e^\theta = 1.4$ clearly outperforms the CTGLR at true value $e^\mu = 1.4$, but the CTCUSUM with $e^\theta = 1.8$ does not clearly outperform the CTGLR chart at $e^\mu = 1.8$ with respect to run length. We can see that the CTGLR has a slightly shorter ARL (by 2 days), but has a bigger SD (by 5 days). From this we conclude that for small values of μ, a CTCUSUM chart with the right value of θ is better than a CT(MAX)GLR chart, while for larger values of μ, the CTGLR chart is likely better. Notice that for $\mu > 2$, the CTGLR outperforms both CTCUSUM charts, although it has a higher variability.

  - From this we conclude that if we know the true out of control rate, the CTCUSUM is preferrable, but if we do not know (especially for large values of μ) or the rate is subject to change over time, the CT(MAX)GLR should be the preferred method.
  - As the CTMAXGLR chart is always larger than the CTGLR chart, it is possible that it would not have such a high variability in detection times due to its earlier detection times.

- The distribution of the run lengths is right-skewed for all charts, as the MRL is always smaller than the ARL. Notably, the distribution of the run lengths for the CTGLR are only very slightly right-skewed, especially for big values of μ.
- The theoretical values only correspond to the Monte Carlo values for big values of μ for the CTCUSUM charts, while for the CTGLR charts the theoretical values are reasonable at all values of μ. This is likely due to the fixed value of θ in CUSUM procedures.
  - The theoretical values using Equation (30) are never close to the real values, with the theoretical values being very small compared to the Monte Carlo values. This is a consequence of a difference in assumptions, as we do not assume the process to start in equilibrium, as well as no longer limiting the CTCUSUM to $C = 1$, opting to consider all failures as qualifying ($C = \infty$).
- For small values of μ, the standard deviation of the run lengths is smaller for the CTGLR than for the CTCUSUM charts, while for bigger values of μ, the SD of the CTCUSUM charts seems to converge towards zero faster than that of the CTGLR charts. This could be due to the fact that the CTGLR adjusts the value of $\hat{\theta}_t$, meaning that higher failure rates also cause the chart to decrease more rapidly, while the CTCUSUM charts have a steady rate of increase and decrease, resulting in more stable detection times.

7.4 Asymptotic study

In section 6.8 we determined some bounds for the asymptotic value of the CTGLR chart. In this section we will assess whether these bounds are reasonable in practice by means of simulation, using the parameters stated in Table 2.

7.4.1 Risk-adjustment

An important property of the continuous time charts is that due to the Proportional Hazards assumption, we are no longer interested in the risk-adjustment terms when performing an asymptotic simulation study ($t \to \infty$). This is due to the fact that the null-rate is also risk adjusted, and we assume the risk-adjustment to be equal for both null-hypothesis as well as alternative hypothesis. Mathematically, this can be shown as follows. Suppose we have n individuals with individual covariates Z_i. Then using (89) we can generate risk-adjusted survival times X_i' and non-risk-adjusted survival times X_i:

$$X_i' = H_0^{-1}[-\log(U_i)e^{-Z_i\beta}] \qquad\qquad X_i = H_0^{-1}[-\log(U_i)]$$

where $U_i \sim U[0,1]$ for all i. Then using (35) with $GLR'(t)$ the risk-adjusted chart and $GLR(t)$ the non-risk-adjusted chart and above generated survival times:

$$\begin{aligned}
&\lim_{t\to\infty} GLR'(t) - GLR(t) \\
&= n\left(-\ln\left(\sum_{i=1}^{n} H_0(X_i')e^{Z_i\beta}\right) + \ln\left(\sum_{i=1}^{n} H_0(X_i)\right)\right) - \sum_{i=1}^{n} H_0(X_i) + \sum_{i=1}^{n} H_0(X_i')e^{Z_i\beta} \\
&= n\left(-\ln\left(\sum_{i=1}^{n} -\ln(U_i)\right) + \ln\left(\sum_{i=1}^{n} -\ln(U_i)\right)\right) - \sum_{i=1}^{n} -\ln(U_i) + \sum_{i=1}^{n} -\ln(U_i) \\
&= 0.
\end{aligned}$$

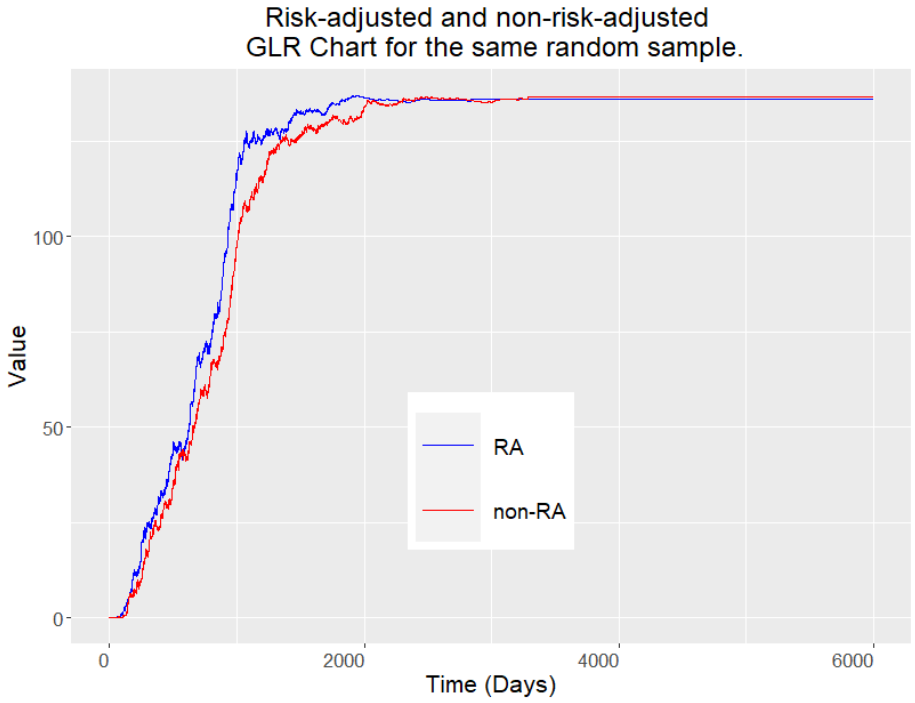


Figure 8: Risk-adjusted and non-risk-adjusted chart for two of the "same" hospital, one having patients with covariates and the other without. The red line indicates the chart when generating outcome times without risk-adjustment and constructing the chart without taking into account risk-adjustment. The blue line is the chart for outcome times generated with risk-adjustment, where we assumed that we perfectly know the corresponding coefficients.

This result can also be seen graphically in Figure 8.

To construct this figure we generated failure times using 2500 randomly generated outcomes, both for the risk- and non-risk-adjusted model, leaving no outcomes censored. The covariates were randomly generated using Bernoulli, Normally and Uniformly distributed random variables. The coefficients were chosen randomly as well. Afterwards the chart was constructed using the corresponding (non-)risk-adjusted baseline hazard, where we assumed that we knew the corresponding coefficients perfectly. We can see that the charts converge towards the same value after all outcomes have been observed. The main conclusion we draw from this is that due to the proportional hazards assumption **any asymptotic result which we obtain for the mean of the non-risk-adjusted charts will also be valid for the risk-adjusted chart, assuming that the regression coefficients can be recovered perfectly.** This statement does not imply that risk-adjustment is useless for this method, it only indicates that assuming we know the risk-adjustment coefficients perfectly, the asymptotic behaviour of the chart is the same whether we generate covariates for the data or not.

Note that this property is not necessarily a limitation of this method, as the risk-adjustment under the null-hypothesis is the risk which we are willing to tolerate. Of course, over time these risk-adjustment factors might change due to a number of outside influences, but this is not important for our simulation study. Therefore we will perform asymptotic analyses under the assumption that the risk-adjustment coefficients can be determined perfectly, as we are not interested in scrutinizing the performance of the Cox Proportional Hazards regression model, but instead are interested in properties of the charts described above. When considering the chart non asymptotically, the risk-adjustment does play a role in the value of the chart, as the failure times under risk adjustment differ from the

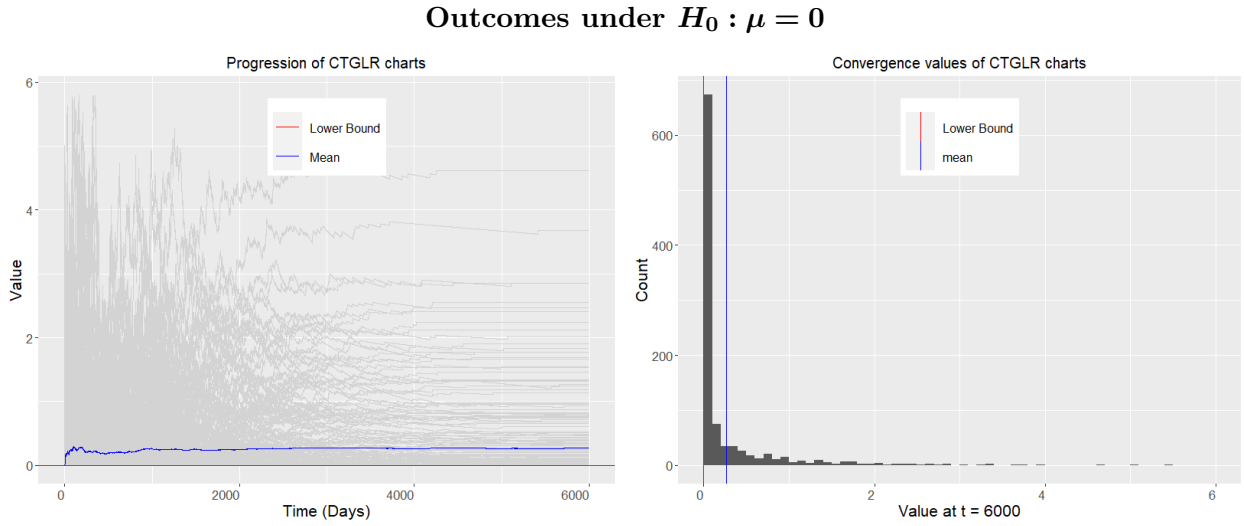
non risk-adjusted failure times.

7.4.2 Generating covariates

If we would like to generate covariates anyway, note that it does not matter how these covariates are generated, as long as we can perfectly recover the coefficients associated with the regression model. The influence of predicting the coefficients wrongly is as follows: if we predict a smaller value than the true one, this means that we will underestimate the risk associated with said covariate. This means that the chart will rise more than it should, resulting in an increased probability of a false alarm. Conversely, if we overestimate a coefficient, the chart will have a longer out of control run length.

7.4.3 Asymptotic behaviour

To evaluate our found lower bound for the GLR chart as $t \to \infty$ in Lemma 6.8.1, we generate 1000 institutions according to Table 2. The time of failure is then generated for every individual using equation (89), assuming no censored observations and $e^{\mu} = 1$. The results of these simulations can be seen in Figure 9.



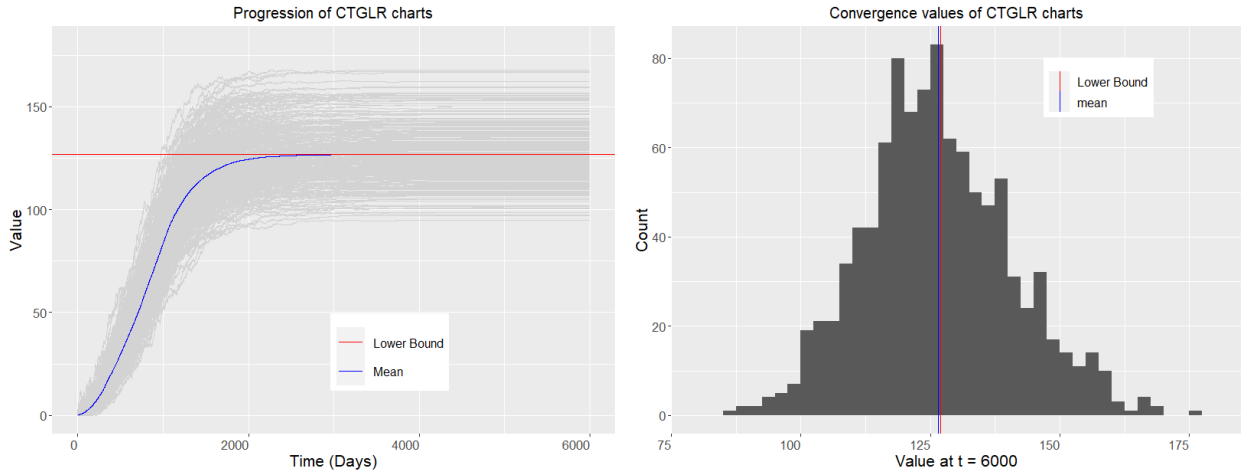
(a) Progression (b) Convergence

Figure 9: On the left 300 CTGLR charts are drawn where outcomes were generated under the null hypothesis. On the right the convergence values (at time 6000) for 1000 of such charts are shown in a histogram. Both graphs have the lower bound as found in equation (50) added in red and the mean value in blue. Every chart represents an institution with 2500 primary procedures in the first 1000 time points.

In this case, the chart always converges above the lower bound, as this bound is equal to zero under the null hypothesis. The upper bound was not added, as it was not very sharp. We can see that only relatively few charts converge towards a value greater than 1 and that more than half of the values are smaller than 0.1. Notable is that some charts have quite a big peak in the beginning, which are caused by very early failures. Even though these charts seem to be very unstable at the beginning, their final path looks to be quite stable. This can be attributed to the fact that the MLE needs quite a few observations before it can accurately approximate the true value of θ. Due to this the CTGLR does not give a very reliable representation at early time points.

Now let us consider the chart for data generated under the alternative hypothesis $e^{\mu} = 1.4$. Once again we generate 2500 procedures uniformly distributed over the first 1000 time points and construct the charts until time 6000. The results can be seen in Figure 10.

Outcomes under $H_1 : \mu = \ln(1.4)$



(a) Progression (b) Convergence

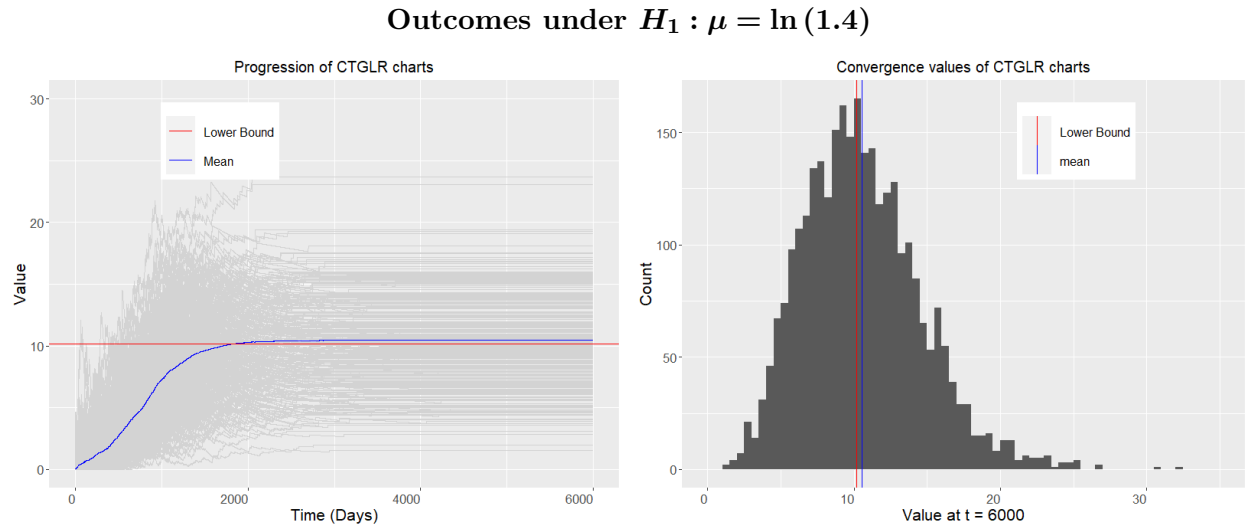
Figure 10: On the left 300 CTGLR charts are drawn for outcomes generated under the alternative hypothesis $\mu = \ln(1.4)$. On the right the convergence values (at time 6000) for 1000 of such charts are shown in a histogram. Both graphs have the lower bound as found in equation (50) added in red and the mean value in blue. Every chart represents an institution with 2500 primary procedures in the first 1000 time points.

We can see that the mean converges towards the lower bound, and does not surpass the lower bound as can be seen in the figure on the right side. The upper bound was not added, as it was again very large. For the exponential distribution we thus have the conjecture that the lower bound can tell us which value we expect the chart to converge to. This means that in theory the CTGLR chart can be used to determine approximately the true value of μ for a set of outcomes which have been sampled, if we know the time to failure for every single individual. It is also possible to obtain approximate confidence intervals (for example, by using bootstrap) for the asymptotic value of the mean.

Notable is also how the mean of the chart keeps rising during the first 1000 time points and starts rising slightly slower around $t = 1000$ until it converges at about $t = 3000$. It is not surprising to see that the mean starts to converge around $t = 3000$ as we saw in Figure 3 that almost all individuals should have already failed about 2000 time points after their initial entry in the study, which is smaller or equal to $t = 1000$ in the graphs above. The graph keeps rising in the beginning because we observe many individuals all at once, which all have a failure time distributed according to the alternative hypothesis, and thus they fail faster than expected.

However, it is likely that when we do not have as many observations the chart will not exhibit this behaviour. Let us determine the behaviour of the chart in the case that we have 200 procedures in the first 1000 time points. The results hereof can be found in Figure 11.

Surprisingly we can see the same behaviour even when we have less observations in the same time frame. Also notable is that the asymptotic distribution for the CTGLR chart seems to be slightly right skewed in all of above figures. We can see that even for 200 observations none of the charts converged towards a value of zero, which is a very desirable property.



(a) Progression (b) Convergence

Figure 11: On the left 300 CTGLR charts are drawn for outcomes generated under the alternative hypothesis $\mu = \ln(1.4)$. On the right the convergence values (at time 6000) for 3000 of such charts are shown in a histogram. Both graphs have the lower bound as found in equation (50) added in red and the mean value in blue. Every chart represents an institution with 200 primary procedures in the first 1000 time points.

8 Discussion

The main methodological contribution of this thesis is the development of the continuous time (maximized) likelihood ratio charts. The main theoretical result is the asymptotic distribution of these charts (Theorem 6.5.1). From this an approximate (as the amount of observations becomes large) average run length for the CTGLR, CTCUSUM and CTMAXGLR charts was derived. Afterwards, this result was evaluated by means of simulation and found to work well, even for detections with small amounts of observations.

Moreover, the simulation results revealed that the CT(MAX)GLR chart has a delay in signalling compared to the CTCUSUM with the correct value of θ. We found that as μ becomes large, the delay becomes smaller. This is similar to the property found by Lorden [25]. Lorden found that the ARL of the discrete time GLR is larger than the ARL of the discrete time CUSUM by a factor proportional to the in control run length divided by the Kullback-Leibler divergence of the null and alternative hypothesis. As μ becomes larger, the KL divergence increases therefore we expect that a similar result can be shown for the continuous time variant. Additionally, we surmise that it it possible to show optimality properties similar to the one shown for the discrete time CUSUM by Lorden [22] and Moustakides [23] for the Exponential family. As the continuous time CUSUM and GLR charts employ totally different assumptions, this could prove to be a challenge. In the simulation results it was apparent that the CTGLR chart had more variability in its detection times than the CTCUSUM chart. Future research could be done to determine the approximate distribution of the average run length of both charts to give more insight what the cost is of not specifying a value of θ.

Simulation wise, the main challenge was the computational intensity of the CTMAXGLR chart. This was partially circumvented by connecting the out of control ARL of the CTGLR with that of the CTMAXGLR in Proposition 6.9.1. It was however very difficult to obtain reliable values for the control limit of the chart, as this required applying the CTMAXGLR chart to in control observations,

forcing a small sample size ($N = 120$) for the estimation of the optimal value of the control limit. As there is no recursive form for the CTMAXGLR, theoretical results for an in control run length are likely difficult to obtain. Future simulation research should be done to determine what realistic gain in detection time could be achieved using the CTMAXGLR instead of the CTCUSUM. Moreover, it is of interest whether the CTMAXGLR has more variance in its detection times compared to the CTCUSUM, as this was one of the drawbacks when comparing the CTGLR with the CTCUSUM.

Interpreting our theoretical results in a medical context, we found that the CTMAXGLR could lead to faster detection times when the true value μ is not known in advance, or when the rate of failure under the alternative is variable over time. Additions to the chart should also be considered. Begun et al. [8] have developed a chart similar to the CTCUSUM with Frailty terms and competing risks. Moreover, procedures after a detection has taken place should be considered, such as setting the chart to a value different than zero aimed at reducing time to a consequent detection. Lucas & Crosier [30] proposed to set the discrete time CUSUM to $h/2$ for Normally distributed data in the discrete time CUSUM, leading to quicker detection times and approximately equal in control run lengths. Finally, it is important to note that the average run length of all considered charts depends on the rate of arrivals ψ. Parameters of the charts should thus be determined according to the size of the hospital in question.

Finally, as this thesis was focussed on comparing Bernoulli CUSUM charts with continuous time CUSUM/GLR charts, we ignored other available continuous time charts for survival data. Future (simulation) research should be done to compare the CTCUSUM and CTMAXGLR with other continuous time monitoring schemes such as the RAST CUSUM by Sego et al. [31], uEWMA for survival time data by Steiner & Jones [32] and finally the STRAND chart by Grigg [33]. Grigg argues that the biggest drawback of the CTCUSUM is the absence of a shuffling mechanism for patients and absence of weighted observations. Consequently, clusters of failures could be considered as noise leading to a delay in detections as well as ignoring the fact that more recent observations are of more interest to detection. Both these issues are addressed in the uEWMA and STRAND charts.

Based on the theoretical considerations we expect that the continuous time charts developed and studied in this part of the thesis could be of value in medical applications. In Part II, we will assess the added value of continuous time charts on a data set consisting of information on transplantation surgery procedures.

Part II
From theory to practice: guidelines for arthroplasty registers

Part I was focussed on the theoretical development of continuous time inspection schemes. In this second part, we will primarily be interested in a practical application of suitable inspection methods to a data set from the Dutch Arthroplasty Register (LROI) [1]. To this end we consider existing methods called funnel plots and Bernoulli CUSUM charts, which will be introduced in sections 9 and 10 respectively. To continue on this, we also consider the continuous time CUSUM chart developed by Biswas & Kalbfleisch [2], which is similar to the Bernoulli CUSUM chart, but processes the available data in real-time. Finally, we consider a method called the CT(MAX)GLR, which was developed by us in section 6. Both these methods will be introduced in section 11. The first three sections of this part therefore introduce the relevant theory behind the models.

Having introduced the reader to the necessary theory we then apply these methods to a data set from the Dutch Arthoplasty Register (LROI) [1] in section 12, in the hope of improving detection speed. The data contains information on all primary hip transplantation surgeries performed in 97 hospitals across the Netherlands, including characteristics of patients such as age, body mass index and diagnosis (see Table 4). The outcome of interest is the revision of the prosthesis. Revision means that the prosthesis has to be replaced, requiring additional transplantation surgery to be performed on the patient. As prostheses are not eternal, every implant has to be revised after a certain amount of time. The date of revision is given in the data set, as well as information on whether and when the outcome was censored (not observed due to some other reason). The most prominent reason for a censored observation in this data set is the death of an individual. The data set is not publicly available, but can be requested from the LROI [1] for research purposes.

Having applied all methods to the data set, we compare them on detection speed, false detection rate and sensitivity. An important problem with the data set is that we do not have any information on whether a hospital had problems in their quality of care at some point in the considered time. Because of this, we perform some additional simulations in section 13 to determine how the obtained results should be interpreted in practice. For this we extrapolate some parameters from the available data set.

9 Funnel plots

The Dutch Arthroplasty Register [1] currently employs funnel plots to detect hospitals with problems in their quality of care [10]. In this chapter we describe the funnel plot method in short, based on the theory in Spiegelhalter et al. [34]. The goal is to compare the detection times produced by funnel plots with those of other methods introduced in later sections.

9.1 Theory

Let $j = 1, ..., k$ represent a hospital, with every hospital performing care on n_j patients. Suppose for every patient we observe a binary outcome $X_{j,i}$, where:

$$X_{j,i} = \begin{cases} 1 & \text{, if patient } i \text{ at hospital } j \text{ had an undesirable outcome within } C \text{ years} \\ 0 & \text{, if patient } i \text{ at hospital } j \text{ had a desirable outcome within } C \text{ years.} \end{cases}$$

We assume that these outcomes are independent and identically Bernoulli distributed, so that $X_{j,i} \sim \text{Ber}(p_j)$ for all i, with p_j the probability of failure at hospital j. Then we can determine the proportion of failures at hospital j as $\gamma_j = \frac{\sum_{i=1}^{n_j} X_{j,i}}{n_j}$. Now we would like to test whether the level of care in a hospital differs from some (national) average. For this we consider the following hypotheses:

$$H_0 : p_j = p_0 \qquad\qquad H_1 : p_j \neq p_0$$

where p_0 is some desired in control failure rate. As often it is unclear which rate is desirable, p_0 is taken as a national average over all available samples. We can choose to take $H_1 : p_j > p_0$ if we are only interested in detecting hospitals which are performing worse than p_0. Now the Central Limit Theorem tells us that (as n_j becomes large):

$$\gamma_j|_{p_j=p_0} \sim \mathcal{N}\left(p_0, \frac{p_0(1-p_0)}{n_j}\right).$$

Therefore we signal a change in quality with confidence level 2α when:

$$\gamma_j \geq p_0 + \xi_{1-\alpha}\sqrt{\frac{p_0(1-p_0)}{n_j}} \text{ or } \gamma_j \leq p_0 + \xi_{\alpha}\sqrt{\frac{p_0(1-p_0)}{n_j}}$$

where ξ_α is the α-th quantile of the standard normal distribution. It is possible to graphically represent this by plotting above bounds in a figure against the amount of procedures in a hospital. By superimposing the points representing the hospitals in this figure we can then graphically determine which hospitals are performing better or worse than average. Notice that due to the choices in this model we are restricted towards considering binary outcomes, which means that we can only consider failures within a certain time period, and ignore all failures (directly) after this cut-off point. This likely leads to a (significant) loss of information, as well as decreasing the detection speed, as we have to wait C years before reporting the outcome. In later sections we will consider methods which do not have this problem.

9.2 Risk-Adjustment

Sometimes it is desirable to adjust the probability of failure of a person depending on some quantities associated with this person, given in the form of a covariate vector Z_i. This vector could include quantities like age, weight, and other relevant statistics for the procedure performed. Heuristically this necessity can be explained as follows: suppose a hospital which is performing treatments sensitive to a genetic defect is located in a region where this gene is very prominent, then it is natural that the failure rate of this hospital will be higher than average. It is however not fair to conclude that the quality of care is lower than that of other hospitals, where cases have a lower risk of failure.

To adjust for this, we recover the probability of failure p_i for person i with covariates Z_i by performing a logistic regression model:

$$\ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta' Z_i.$$

The probability of failure for person i is then given by:

$$p_i = \frac{1}{1 + e^{-\beta_0 + \beta' Z_i}} \tag{62}$$

so that we can determine the expected amount of failures E_j at hospital j as:

$$E_j = \mathbb{E}[\sum_{i=1}^{n_j} X_{j,i}] = \sum_{i=1}^{n_j} p_i = \sum_{i=1}^{n_j} \frac{1}{1 + e^{-\beta_0 + \beta' Z_i}}.$$

Suppose we observe O_j failures at hospital j. Then we obtain a risk-adjusted proportion of failures as:

$$\gamma_{j,RA} = \frac{O_j}{E_j} \cdot p_0.$$

9.3 Overdispersion

Often it is the case that the assumed null distribution (in this case the Bernoulli distribution) does not completely capture the variance of the observations. This happens mostly due to missing covariates in the risk-adjustment method. To account for this we can determine an **overdispersion** factor, used to correct the variance of the null-distribution. Spiegelhalter et al. [34] propose 2 methods to determine such an overdispersion factor, covered in the following 2 sections.

9.3.1 Multiplicative method

We determine an overdispersion factor $\hat{\phi}$ as:

$$\hat{\phi} = \frac{1}{I} \sum_i \frac{(x_i - p_0)^2}{\frac{p_0(1-p_0)}{n_i}}$$

where I is the number of in control samples and n is the total number of samples and the summation is over the in control samples. There are multiple ways to determine which samples are in control, which we will not discuss here. After determining this factor, we then obtain new confidence bounds by multiplying the variance of the null-distribution by the factor $\hat{\phi}$, which yields the bounds:

$$p_0 \pm \xi_{1-\alpha} \sqrt{\hat{\phi} \cdot \frac{p_0(1-p_0)}{n_j}}.$$

9.3.2 Additive method

We determine an overdispersion factor $\hat{\tau}^2$ as:

$$\hat{\tau}^2 = \frac{I\hat{\phi} - (I-1)}{\sum_i \omega_i - \frac{\sum_i \omega_i^2}{\sum_i \omega_i}}$$

where I is as above, $\omega_i = \frac{1}{s_i^2}$ and $\hat{\phi}$ is a test for heterogeneity, such that $\hat{\tau}^2 = 0$ when $\hat{\phi} < \frac{(I-1)}{I}$. Then we adjust the variance of the null distribution by an additive factor of $\hat{\tau}^2$, so that the bounds become:

$$p_0 \pm \xi_{1-\alpha} \sqrt{\frac{p_0(1-p_0)}{n_j} + \hat{\tau}^2}.$$

Overdispersion methods will not be considered in the rest of this thesis, as these require us to know which samples were in control and which were not. Unfortunately this knowledge is not available to us in the available data set. As the name suggests, a funnel plot is a graphical tool to analyse which

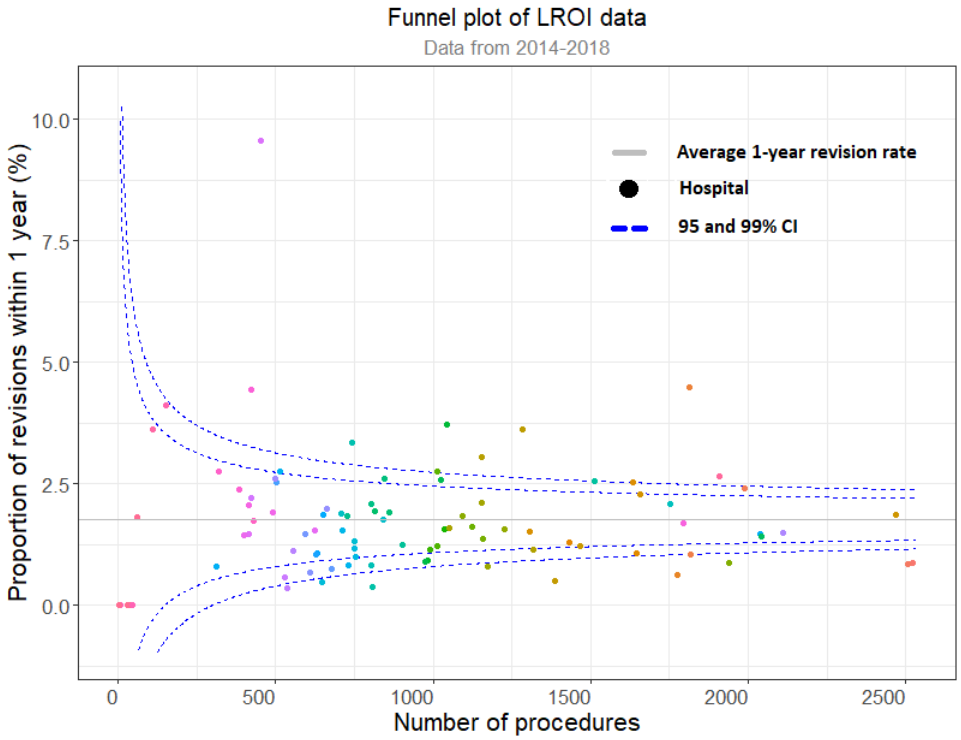


Figure 12: Funnel plot of the first 4 years of data from the LROI [1] with 1−year post operation follow-up. Every dot represents a distinct hospital, the horizontal grey line indicates a national 1−year post transplant rate of failure and the blue lines indicate 95 (inner) and 99 (outer) percent confidence intervals.

hospitals are performing better or worse than average. A risk-adjusted funnel plot for the first 4 years of data from the LROI data set [1] with 1−year follow-up can be seen in Figure 12. All hospitals which are outside of the confidence interval bounds are detected by this method. Hospitals lying above the dotted blue lines are performing worse than expected (at 95 and 99 percent confidence) thus warranting an inspection, and below the blue lines are performing better than expected. We can see that depending on the required confidence level, a different amount of hospitals would be detected.

10 Bernoulli CUSUM charts

In this section we restate the discrete time CUSUM chart as introduced in section 4.2 in the case that the outcomes follow a Bernoulli distribution. The exposition will be based on sections 2.2.2 and 3.1 of the Master's thesis written by Caroline Kok [9]. This section will be written in the context of the LROI [1], where we are interested in detecting a hospital with a decrease (or increase) in the quality of care. Van Schie et al. [11] have already evaluated Bernoulli CUSUM charts on the LROI data set, and concluded that faster detection of a decrease in the quality of care is possible compared to funnel plots (section 9).

10.1 Definition of the Bernoulli CUSUM chart

In contrast to funnel plots, in a Bernoulli CUSUM chart we determine individual charts for each hospital. Suppose we have an influx of patients at the hospital in question, with binary outcomes X_i, with $i = 1, 2, ...$, where:

$$X_i = \begin{cases} 1 & \text{, if patient } i \text{ had an undesirable outcome within } C \text{ years} \\ 0 & \text{, if patient } i \text{ had a desirable outcome within } C \text{ years} \end{cases}$$

where we assume that the outcomes are i.i.d. Bernoulli, such that $X_i \sim \text{Ber}(p)$. The Bernoulli CUSUM chart then tests hypotheses of the form:

$$H_0 : X_1, X_2, ... \sim \text{Ber}(p_0) \tag{63}$$
$$H_1 : X_1, ..., X_{\nu-1} \sim \text{Ber}(p_0) \tag{64}$$
$$X_\nu, X_{\nu+1}, ... \sim \text{Ber}(p_1) \tag{65}$$

with $p_1 > p_0$ and $\nu \in \mathbb{N}_{>0}$. We are therefore trying to determine whether a change of distribution has happened at some time point τ, which we define as the time of entry into the study of patient ν. Notice that using the funnel plot method we tested for $H_1 : p \neq p_0$, where we implicitly chose the value of p_1 through the confidence level α. The Bernoulli CUSUM chart tests whether the probability of failure at the hospital has increased starting from observation ν (similarly: at some time point τ). This is in contrast to funnel plots, where we test whether the observations were failing at a higher rate from the start. Because of this it is likely that the Bernoulli CUSUM chart will yield a more powerful test, as it is not guaranteed that the quality of care has decreased from the start of observations. The Bernoulli CUSUM chart S_n after n observations is then simply the likelihood ratio test associated with these hypotheses.

$$S_n = \max_{1 \leq \nu \leq n} \sum_{i=\nu}^{n} \ln\left(\frac{f_{\theta_1}(Y_i)}{f_{\theta_0}(Y_i)}\right) = \sum_{i=1}^{n} \ln\left(\frac{f_{\theta_1}(Y_i)}{f_{\theta_0}(Y_i)}\right) - \min_{1 \leq \nu \leq n} \sum_{i=1}^{\nu} \ln\left(\frac{f_{\theta_1}(Y_i)}{f_{\theta_0}(Y_i)}\right). \tag{66}$$

We then **reject the null hypothesis** at observation n when $S_n \geq h$, where $h > 0$ is called the **control limit**. Alternatively, it is possible to rewrite this using $W_n = \ln\left(\frac{f_{\theta_1}(Y_n)}{f_{\theta_0}(Y_n)}\right)$ to obtain:

$$S_n = \max\left(0, S_{n-1} + W_n\right). \tag{67}$$

Substituting the Bernoulli distribution into above expression we obtain:

$$W_n = \ln\left(\frac{p_1^{X_n}(1-p_1)^{1-X_n}}{p_0^{X_n}(1-p_0)^{1-X_n}}\right) = X_n \ln\left(\frac{p_1(1-p_0)}{p_0(1-p_1)}\right) + \ln\left(\frac{1-p_1}{1-p_0}\right). \tag{68}$$

Note that the Bernoulli CUSUM chart is simply a likelihood ratio test of the associated hypotheses, where the first $\nu - 1$ observations are ignored, with ν determined such that the value of the chart is maximal. This is due to the fact that the first $\nu - 1$ observations are assumed to be in control. Let $N_1 = \inf\{n : \sum_{i=1}^{n} W_i \geq h\}$, it was shown in section 4.2 that the average run length of a discrete time CUSUM chart is then given by:

$$\mathbb{E}_i[\tau] = \frac{\mathbb{E}_i[N_1]}{\mathbb{P}_i(\sum_{i=1}^{N_1} W_i \geq h)}$$

where $i = 0, 1$ indicates the null and alternative hypothesis.

10.2 Alternative formulation

It is sometimes convenient to formulate the problem in 10.1 in a different way, so that it becomes comparable to the charts considered below. Instead of thinking about failure probabilities we define the **odds-ratio** under the alternative distribution as:

$$OR = \frac{\frac{p_1}{1-p_1}}{\frac{p_0}{1-p_0}} = \frac{p_1(1-p_0)}{p_0(1-p_1)} =: e^{\theta}. \tag{69}$$

Then the hypothesis stated in (63) become:

$$H_0 : OR = 1 \qquad\qquad H_1 : OR = OR_A \tag{70}$$

where we choose $OR_A = e^{\theta} > 1$. We are therefore testing whether the **odds** $\frac{p_1}{1-p_1}$ differ by a factor of e^{θ} from the null-odds $\frac{p_0}{1-p_0}$. Using some algebra we easily recover p_1 from p_0 and θ as follows:

$$p_1 = \frac{p_0 e^{\theta}}{(1-p_0)(1+p_0 e^{\theta})}. \tag{71}$$

The Bernoulli CUSUM chart S_n is then defined exactly as in (67), where we can rewrite W_n in (68) as:

$$W_n = X_n \ln(OR_A) + \ln\left(\frac{1}{1 - p_0 + OR_A \cdot p_0}\right) \tag{72}$$

so that we can replace the value of p_1 with the associated value of θ.

10.3 Risk-adjustment

When we know some characteristics of the patients at the hospital it is sometimes desirable to adjust for the risk incurred by these characteristics. These characteristics are denoted by a covariate vector Z_i for person i. Similarly to section 9.2 we can then recover probabilities of failure for person i using logistic regression under both hypotheses:

$$p_0^i = \frac{1}{1 + e^{-\beta_0 + \beta' Z_i}} \qquad\qquad p_1^i = \frac{1}{1 + e^{-(\theta + \beta_0 + \beta' Z_i)}}$$

where we assume that the risk-adjustment coefficients are equal under both hypothesis, but the intercept term differs by a factor of θ under the alternative hypothesis. We can then determine the odds-ratio under the alternative hypothesis:

$$OR_A = \frac{p_1^i(1-p_0^i)}{p_0^i(1-p_1^i)} = e^{\theta}$$

which does not depend on the risk adjustment terms. Then the risk-adjusted Bernoulli CUSUM chart $S_{n,RA}$ tests the hypotheses:

$$H_0 : OR = 1 \qquad\qquad H_1 : OR = OR_A = e^{\theta}$$

for some $e^{\theta} > 1$, which is exactly the same as in (70). The risk-adjusted CUSUM chart $S_{n,RA}$ is then defined by:

$$S_{n,RA} = \max\left(0, S_{n-1,RA} + W_{n,RA}\right)$$

where:

$$W_{n,RA} = X_n \ln\left(OR_A\right) + \ln\left(\frac{1}{1 - p_0^n + OR_A \cdot p_0^n}\right).$$

An example of a risk-adjusted Bernoulli CUSUM chart can be seen in Figure 13. In this figure the Bernoulli CUSUM chart is plotted for a hospital from the LROI data set [1]. A value of $e^{\theta} = 2$ is chosen, as well as a control limit of $h = 10$. We can see that the null-hypothesis is rejected around $t = 1350$, as the chart has surpassed the control limit. Note how the chart only starts at time $t = 365$, as our first outcome is only observed $1-$year post beginning of the study. It is possible to plot the Bernoulli CUSUM chart before this time as well, but obviously it would stay at a value of zero in the first year.

The two biggest advantages of the Bernoulli CUSUM chart are that it is very easy to construct and interpret. Besides this, the CUSUM procedure is a sequential test, which is more suited towards the conclusions we want to draw. Unfortunately this procedure suffers from the same drawback as the funnel plot - the outcomes considered have to be dichotomized. This means that patient outcomes are only considered C years after their primary procedure, and information can be lost if the outcome is observed slightly past the C years. To counteract this fact, in the next sections we consider procedures which will integrate all the information available to us about every individual into the chart at any time.

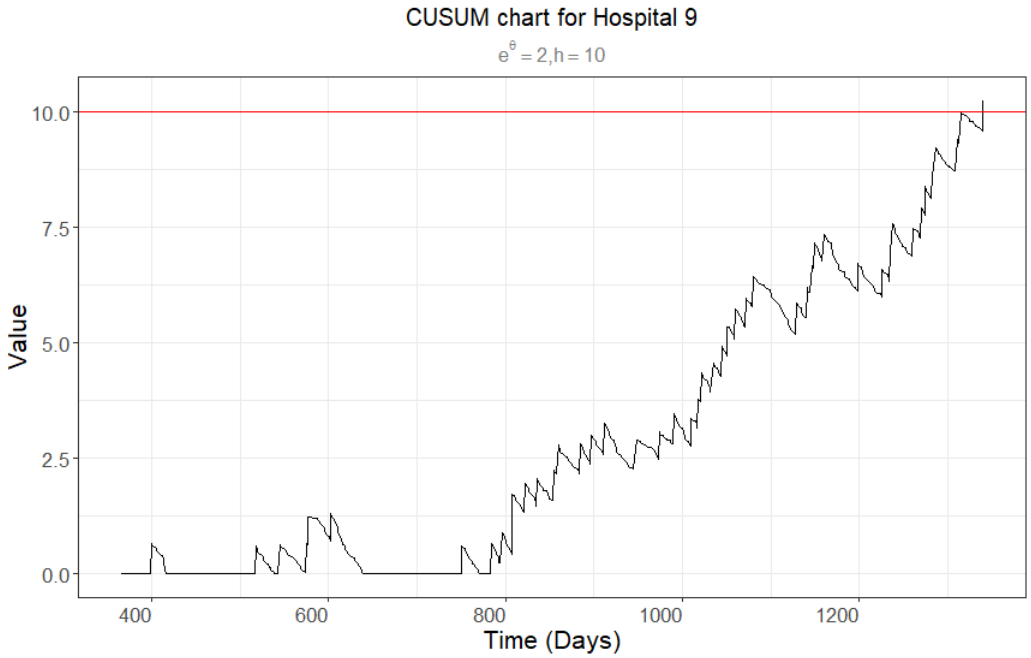


Figure 13: Bernoulli CUSUM chart with 1−year follow up for hospital 9 from the LROI data set [1]. Alternative hypothesis value $e^{\theta} = 2$ and control limit $h = 10$ were chosen. The control limit is shown using a horizontal red line. The null hypothesis is rejected around $t = 1350$ as the chart crosses the control limit.

11 Continuous time control charts

In the previous sections we have introduced two methods which use a discrete time (Bernoulli) outcome to (non-sequentially and sequentially) test for a decrease in the quality of care at a hospital. As the data set available to us contains survival outcomes and is therefore of longitudinal nature, we are interested in some sequential tests which can fully utilise all the information available to us at any time point. This is possible by also integrating the information about a patient being alive into the chart, instead of only considering a patient a certain amount of years post procedure. For this reason, we will introduce three methods which combine our knowledge of survival analysis with likelihood ratio tests, in the hope of improving detection speed.

The following sections will all be structured in the same way. First we restate the derivation of the charts, which were already introduced in sections 5, 6 and 6.9. Afterwards, we state some key properties which were shown in these same chapters, without delving into the proof. This way we can apply the charts to our data set and assess our findings using the available theory.

11.1 Notation

In the following sections we will adhere to the notation in Biswas & Kalbfleisch [2]. We consider a hospital performing primary hip replacement surgery. Let X_i denote the time from the primary procedure to the time of revision for patient i. Then define S_i as the time of primary procedure from some starting point of the study. The chronological time of failure is then given by $T_i = S_i + X_i$. Let C_i denote the chronological censoring time of the patient, if applicable. Additionally, for every patient we have covariates, which are denoted by the $p-$vector Z_i. The covariates for all patients can then be combined row-wise to form the matrix Z. Additionally, we assume that there is a known (risk-adjusted)

null-distribution for X_i, defined by the hazard function $h_i(x)$ (note that x here is the time to revision). For $\mu > 0$ define $h_i^\mu(x) = h_i(x) \cdot e^\mu$. Let this notation also carry over to the cumulative hazard rate $H_i^\mu(x)$ as well as the survival function $S_i^\mu(x)$ and the density function $f_i^\mu(x)$. With this notation the subscript indicates the risk-adjustment term for person i and the superscript denotes the factor e^θ by which the hazard differs from the null-rate (which is indicated without superscript). A characteristic without superscript indicates that $\mu = 0$, so that the notation looks neater. Additionally, we denote by $h_0(x)$ the non risk-adjusted hazard rate under the null hypothesis (there is no 0-th person thus this notation is consistent).

Now define $N^A(t) = \sum_{i \geq 1} \mathbb{1}_{\{S_i \leq t\}}$ to be the number of primary procedures (transplants) in $[0, t]$. Define $\tilde{N}_i^D(t) = \mathbb{1}_{\{T_i \leq t\}}$ as a failure indicator for patient i up to time t. Define, for some constant $C > 0$:

$$Y_i(t) = \mathbb{1}_{\{S_i \leq t \leq S_i + C \cap T_i \cap C_i\}}$$

as an indicator whether patient i is **active**. If a person is not active, we call them **inactive**. This means that people are only active after they have had a primary procedure and only up until the point that they have either failed, been censored or reached C years post transplant. Define $N_i^D(t) = \int_0^t Y_i(u) d\tilde{N}_i^D(u)$ for $t > 0$ as the counting process for a qualifying failure of patient i. Finally, define the **cumulative intensity** (see (4)):

$$\Lambda_i^\mu(t) = \int_0^t Y_i(u) \cdot h_i^\mu(u) du \tag{73}$$

with subscript and superscript as above. Note that for indices i which have not yet had a primary procedure, so when $t < S_i$, we have that $\Lambda_i^\mu(t) = 0$, due to the definition of $Y_i(t)$. Some characteristics of commonly used survival distributions can be found in Table 1.

11.2 CTCUSUM

Biswas & Kalbfleisch [2] developed a continuous time CUSUM (CTCUSUM) procedure using the Cox proportional hazards model (see section 2.6.1). In this section we introduce their method, with a slight change: in their article Biswas & Kalbfleisch consider outcomes of patients only up until 1 year post transplant. We will generalize this assumption to considering patients up until C years post transplant. Remember that the Cox regression model is given by:

$$h_i(x) = h_0(x) e^{Z_i^\top \beta} \text{ for } x > 0.$$

Now using the chronological time t (from the start of the study) define a counting process $N_i^D(t)$ corresponding to the $i-$th patient as above. We then have:

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

Now denote $dH_i(t) = h_i(t - S_i)dt$, and note that this represents the instantaneous hazard of a revision. The term e^μ is the factor by which the hazard at an institution differs from the national rate $dH_i(t)$, or equally the null rate with $\mu = 0$.

We want to calculate a likelihood ratio statistic corresponding to a test of $\mu = 0$ versus $\mu = \theta$ ($\theta > 0$ known), therefore testing whether the quality of transplantations has decreased. Consider

the likelihood based on $dN_i^D(t)$, the response in the interval $(t, t+dt]$ conditional on the information available up to time t. We obtain the likelihood ratio in this interval using equation (5):

$$\begin{aligned} LR'_\theta(t, t+dt) &= \frac{\prod_{i\geq 1} \left(e^\theta d\Lambda_i(t)\right)^{dN_i^D(t)} \left(1 - e^\theta d\Lambda_i(t)\right)^{1-dN_i^D(t)}}{\prod_{i\geq 1} \left(e^0 d\Lambda_i(t)\right)^{dN_i^D(t)} \left(1 - e^0 d\Lambda_i(t)\right)^{1-dN_i^D(t)}} \\ &= \prod_{i\geq 1} \frac{\left(e^\theta\right)^{dN_i^D(t)} \left(1 - e^\theta d\Lambda_i(t)\right)^{1-dN_i^D(t)}}{\left(1 - e^0 d\Lambda_i(t)\right)^{1-dN_i^D(t)}}. \end{aligned}$$

Remember that individuals stop providing information to the chart after they are censored, but the information acquired until the time of censoring is taken into regard. The implications of this construction will be discussed below. Now using a repeated conditioning argument, the likelihood based on the information up to time t can be calculated using equation (6) and is equal to:

$$LR_\theta(t) = \prod_{i\geq 1} \frac{\left(e^\theta\right)^{N_i^D(t)} e^{-e^\theta \Lambda_i(t)}}{e^{-\Lambda_i(t)}}.$$

As a consequence, the log likelihood ratio up to time t is given by:

$$\begin{aligned} U_\theta(t) := \ln(LR_\theta(t)) &= \sum_{i\geq 1} \ln\left(\frac{\left(e^\theta\right)^{N_i^D(t)} e^{-e^\theta \Lambda_i(t)}}{e^{-\Lambda_i(t)}}\right) && (74) \\ &= \sum_{i\geq 1} \theta N_i^D(t) + \left(-e^\theta \Lambda_i(t) + \Lambda_i(t)\right) && (75) \\ &= \theta N^D(t) - (e^\theta - 1) \sum_{i\geq 1} \Lambda_i(t) && (76) \end{aligned}$$

Because we would like to test the hypothesis of a change of hazard rate starting from some patient τ, the CTCUSUM chart has a cut-off at zero using a similar argument as in section 4.2. To continuously update the chart we are interested in the increments dUt defined as:

$$dU_\theta(t) = \theta dN^D(t) - (e^\theta - 1) \sum_{i=1}^{n} d\Lambda_i(t). \tag{77}$$

The (one-sided) continuous time CUSUM chart is then given by:

$$G_\theta(t+dt) = \max(0, G_\theta(t) + dU_\theta(t)) \text{ for } t > 0 \tag{78}$$

which is equivalent to:

$$G_\theta(t) = U_\theta(t) - \min_{0\leq s\leq t} U_\theta(s) \text{ for } t > 0. \tag{79}$$

Note that if we wanted to test whether there was an increase in quality (therefore decrease in the amount of revisions) we could plot:

$$G_\theta^-(t) = \min(0, G_\theta^-(t) - dU_\theta(t)) \text{ for } t > 0.$$

We will not pursue this further in this thesis. Notice how equations (78) and (79) are the continuous analogues of equations (67) and (66).
We conclude this section by summarising the chart in the following definition:

Definition 11.2.1. *The **continuous time cumulative sum chart (CTCUSUM)** is given by:*

$$G_\theta(t) = \max\left(0, \theta N^D(t) - (e^\theta - 1)\sum_{i\geq 1}\Lambda_i(t)\right). \tag{80}$$

This chart is used to test the hypotheses of a change in cumulative intensity starting from patient ν:

$$\begin{aligned} H_0 :& X_1, X_2, ... \sim \Lambda_i \\ H_1 :& X_1, X_2, ... \sim \Lambda_i \\ & X_\nu, X_{\nu+1}, ... \sim \Lambda_i^\theta \end{aligned}$$

*for $\nu \geq 1$ unknown and $\Lambda_i^\theta = e^\theta \cdot \Lambda_i$ the risk-adjusted baseline hazard rate multiplied by e^θ and $\theta > 0$ a constant chosen in advance. The null hypothesis is rejected at time t when $G_\theta(t) \geq h$ for some $h > 0$, called the **control limit**.*

Note that the CTCUSUM and Bernoulli CUSUM from section 10 test the same hypothesis, but have a different underlying model. The Bernoulli CUSUM assumes that outcomes are Bernoulli distributed, while the procedure considered in this section assumes that the distribution follows the Cox proportional hazard model. The major advantage of using a continuous time version of the CUSUM is that now all outcomes can be considered, irrespective of their failure time past primary procedure. Besides this, the CTCUSUM also incorporates the information about the survival of a patient up to every time point t. This can be seen in equation (80), where the chart rises by a fixed value of $\theta > 0$ whenever a failure is observed, and drifts downwards at every time-point by $e^\theta - 1$ (called the excess risk) multiplied by the accumulated cumulative baseline hazard rate $\sum_{i\geq 1}\Lambda_i(t)$ of all active patients. It was shown in section 6.8 that under the null-hypothesis an individual is expected to fail when their cumulative hazard reaches a value of one: $\Lambda_i(t) = 1$. This means that under the null-hypothesis, over his/her lifetime every patient will contribute to an increase in the chart of θ, and a decrease of $e^\theta - 1$. It is trivial to show that $e^\theta - 1 \geq \theta$, therefore under the null-hypothesis every individual will have a negative influence on the chart, which means that the chart is not expected to significantly rise above a value of zero at any time under the null hypothesis as the CTCUSUM has a cut-off at zero.

11.2.1 Properties

We summarise some of the key properties of the CTCUSUM stated in sections 5 and 6.

- The CTCUSUM chart is non-negative and can only jump upwards by a fixed amount θ.
- The CTCUSUM is used to test the hypothesis of a decrease in quality of care starting from observation ν, or similarly starting from time τ. The null hypothesis is rejected when the chart surpasses the value of h, called the control limit.
- Assuming observations start from time $t = -C$ and that the process thus starts at equilibrium at $t = 0$ Biswas & Kalbfleisch [2] obtained the following approximation for the average run length of the CTCUSUM chart:

$$\mathbb{E}[\tau_h] = \begin{cases} \frac{h}{\eta} - \frac{e^{-\mu}(e^\theta - 1)}{\eta}\left(\frac{1-e^{-\omega_0\theta}}{1-e^{-\omega_0 h}}\right), & \eta \neq 0 \\ \frac{h^2 e^{-\mu}}{\theta^2\gamma}, & \eta = 0 \end{cases} \tag{81}$$

  with h the control limit, $\eta = (\theta e^\mu - e^\theta + 1)\gamma$, ω_0 the solution to $\omega_0(e^\theta - 1) + e^\mu(e^{-\omega_0\theta} - 1) = 0$ and γ as in equation (26). In practice this expression for the average run length is not very useful, as the two assumptions require us to ignore the first C years of observations.

- An approximate asymptotic average run length for the CTCUSUM chart with $C = \infty$ was determined in Corollary 6.10.1, denoted by $ARL(\theta, \mu, h)$ which is the solution to the equation:

$$\mathbb{E}[G_\theta(t)] = \left(\theta + e^{-\mu} - e^{\theta}e^{\mu}\right) I(\mu, t) = h \tag{82}$$

  where $\theta > 0$ as in Definition 11.2.1, μ the true value of failure at the hospital in question and $I(\mu, t)$ as defined in Lemma 6.4.4.

11.3 CTGLR

In this section we will restate the derivation of the continuous time maximized generalized likelihood ratio chart (CTGLR) from section 6. Similarly to the CTCUSUM in the previous section, the CTGLR uses a Cox proportional hazards model to determine a likelihood ratio test statistic. The CTGLR however is used for testing whether a hospital has a failure rate higher than some (in control) rate starting from **the beginning of observations**. This means that the CTGLR is used to test a different hypothesis than the CTCUSUM chart. Besides this, with the CTCUSUM chart we assumed to know the value of θ indicating the increased failure rate with respect to the baseline. The CTGLR will no longer require this assumption, making it a more generally applicable method. Finally, the CTGLR can be extended to test the same hypothesis as the CTCUSUM chart, which will be called the CTMAXGLR chart and derived in section 11.4. The CTMAXGLR however is very difficult to determine properties of, therefore we will derive some properties of the CTGLR chart, which are then easily generalized to hold for the CTMAXGLR chart.

11.3.1 Notation

The notation will stay unchanged from section 11.1, except that instead of $Y_i(t) = \mathbb{1}_{\{S_i \leq t \leq S_i + C \cap T_i \cap C_i\}}$ we will now define

$$Y_i(t) = \mathbb{1}_{\{S_i \leq t \leq T_i \cap C_i\}}$$

as an indicator whether patient i is **active**. If a person is not active, we call them **inactive**. This means that people are only active after they have had a primary procedure and only up until the point that they have either failed or been censored, meaning we no longer restrict ourselves to C years post transplant. We remind the reader that the **cumulative intensity** is given by:

$$\Lambda_i^\mu(t) = \int_0^t Y_i(u) \cdot h_i^\mu(u) du \tag{83}$$

with $\Lambda_i(t) := \Lambda_i^0(t)$. Note that for indices i which have not yet had a primary procedure, so when $t < S_i$, we have that $\Lambda_i^\mu(t) = 0$, due to the definition of $Y_i(t)$. A table with characteristics for some commonly used survival distributions can be found in Table 1.

11.3.2 Derivation

We will consider a generalized likelihood ratio chart using a proportional hazards assumption for the outcome distribution, based on the theory in section 5. Similarly, we also consider the counting processes as defined in section 5.2 where we defined a counting process $N_i^D(t)$ corresponding to the $i-$th patient such that:

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

The difference will lie in the alternative hypothesis considered. Let e^{μ} be the true factor by which the hazard at the hospital is higher than the baseline hazard. This time we want to calculate a likelihood ratio statistic corresponding to a test of $\mu = 0$ versus $\mu = \theta$ for some **unknown** $\theta > 0$, again testing whether the quality of transplantations has decreased. For the continuous time CUSUM we assumed θ to be known. The likelihood ratio based on $dN_i^D(t)$, using the information from the response in the interval $(t, t + dt]$ conditional on the information available up to time t is then given by (see equation (5)):

$$\begin{aligned} LR'_{\hat{\theta}_t}(t) = LR'_{GLR}(t) &:= \sup_{\theta>0} \frac{\prod_{i\geq 1} \left(e^{\theta} d\Lambda_i(t)\right)^{dN_i^D(t)} \left(1 - e^{\theta} d\Lambda_i(t)\right)^{1-dN_i^D(t)}}{\prod_{i\geq 1} \left(e^{0} d\Lambda_i(t)\right)^{dN_i^D(t)} \left(1 - e^{0} dH_i(t)\right)^{1-dN_i^D(t)}} \\ &= \prod_{i\geq 1} \frac{\left(e^{\hat{\theta}_t}\right)^{dN_i^D(t)} \left(1 - e^{\hat{\theta}_t} d\Lambda_i(t)\right)^{1-dN_i^D(t)}}{\left(1 - d\Lambda_i(t)\right)^{1-dN_i^D(t)}} \end{aligned}$$

where $\hat{\theta}_t$ is the MLE over $\theta > 0$ at time t. We will discuss how to determine this estimate in 11.3.1. Once again using a repeated conditioning argument, the likelihood based on the information up to time t can be calculated using equation (6) and is equal to:

$$LR_{GLR}(t) = \prod_{i\geq 1} \frac{\left(e^{\hat{\theta}_t}\right)^{dN_i^D(t)} e^{-e^{\hat{\theta}_t}\Lambda_i(T_i)}}{e^{-\Lambda_i(T_i)}}.$$

Thus the log likelihood ratio up to time t is given by:

$$\ln(LR_{GLR}(t)) = \hat{\theta}_t N^D(t) - (e^{\hat{\theta}_t} - 1) \sum_{i\geq 1} \Lambda_i(t).$$

We can determine the ML estimate of θ at time t, which will be done in the following lemma.

Lemma 11.3.1. *The maximum likelihood estimate of θ for the Generalized likelihood ratio test introduced in section 11.3.2 is given by:*

$$\hat{\theta}_t = \max\left(0, \ln\left(\frac{N_D(t)}{\sum_{i\geq 1} \Lambda_i(t)}\right)\right). \tag{84}$$

Proof. To determine the maximum likelihood of θ, we first remind ourselves that we are determining the maximum likelihood estimator for θ for the hypotheses $\mu = 0$ against $\mu = \theta$ with θ unknown. Heuristically this means that we are testing whether the true *intensity* $e^{\mu} d\Lambda_i(t)$ differs from the national rate $d\Lambda_i(t)$, which was pre-determined from some sort of training set (a data set which we know to have in control procedures). Whenever we experience a failure (revision) we obtain some information as to how likely it was that the individual in question would experience a failure at that time point, this will then change the value of the MLE accordingly. Consider the likelihood up to time t using equation (7):

$$L(\theta|t) = \prod_{i\geq 1} \left(e^{\theta} d\Lambda_i(t)\right)^{dN_i^D(t)} e^{-e^{\theta}\Lambda_i(T_i)}$$

where we have n observations which are either active or have had an outcome or had their outcome censored. The logarithm of the likelihood (up to time t) is then given by:

$$l(\theta) = \sum_{i \geq 1} (dN_i^D(t))(\theta + \ln(d\Lambda_i(t))) - e^\theta \Lambda_i(t).$$

Taking the derivative w.r.t. θ and equating to zero we then obtain:

$$\sum_{i \geq 1} (dN_i^D(t)) - e^\theta \sum_{i \geq 1} \Lambda_i(t) = 0$$

which yields:

$$e^\theta = \frac{N^D(t)}{\sum_{i \geq 1} \Lambda_i(t)}.$$

The MLE for θ at time t is then given by:

$$\hat{\theta}_t = \max\left(0, \ln\left(\frac{N_D(t)}{\sum_{i \geq 1} \Lambda_i(t)}\right)\right)$$

where the cut-off at zero arises from the fact that we test the hypothesis of $\mu = 0$ against a hypothesis of $\mu = \theta > 0$. □

We summarise the results from the previous subsection by defining the Continuous Time Generalized Likelihood Ratio chart:

Definition 11.3.1. *The continuous time generalized likelihood ratio chart (CTGLR) is given by:*

$$GLR(t) = \hat{\theta}_t N^D(t) - (e^{\hat{\theta}_t} - 1) \sum_{i \geq 1} \Lambda_i(t). \tag{85}$$

This chart is used to test the hypothesis that the cumulative intensity at an institution differs by a factor of e^θ (with $\theta > 0$ unknown) from the null intensity $\Lambda_i(t)$, this can be stated as:

$$H_0 : \mu = 0$$
$$H_1 : \mu = \theta.$$

The counting processed are as defined in section 11.1. The maximum likelihood estimator $\hat{\theta}_t$ was found in Lemma 11.3.1:

$$\hat{\theta}_t = \max\left(0, \ln\left(\frac{N^D(t)}{\sum_{i \geq 1} \Lambda_i(t)}\right)\right). \tag{86}$$

*The null hypothesis is rejected at time t when $GLR(t) \geq h$ for some $h > 0$, called the **control limit**.*

The greatest power of the CTGLR is that it is no longer necessary to select a value of $\theta > 0$ a priori. The price paid for this generalization is that it is not possible to recursively determine the value of the CTGLR statistic, in contrast to the CTCUSUM (see equation (78)). This is because the CTGLR requires us to calculate the value of $\hat{\theta}_t$ at every relevant t, and then recalculate the total value of the statistic using the determined MLE. This means that the CTGLR is more computationally intensive than the CTCUSUM, even though it only tests for an immediate change in the rate of failures.

Another important consideration is that with the CTCUSUM the value of the alternative hypothesis is specified through the chosen value of θ, meaning that the control limit h could be chosen in order to optimise a quantity of interest (i.e. ARL under the alternative hypothesis, power in K years etc.). With the CTGLR the interpretation of h is slightly different, as there is no longer a specific value of the alternative hypothesis we are testing for. This means that implicitly, when choosing a value for h, we are choosing which value θ we want to work with. This means that the CTGLR chart becomes harder to interpret for laymen, as the purpose of the control limit is no longer easily related to the purpose of the control limit of the CTCUSUM and Bernoulli CUSUM.

11.3.3 Properties

In this section we state some properties of the CTGLR which were determined in section 6. Many of these properties will later be generalized to also be applicable for the CTMAXGLR chart in section 11.4.1.

- In Lemma 6.3.1 the CTGLR was shown to be non-negative for every $t \geq 0$ and strictly positive whenever $N^D(t) > \sum_{i \geq 1} \Lambda_i(t)$. Similarly to the CTCUSUM, the CTGLR therefore has a "cut-off" at zero. The interpretation of the CTGLR and the CTCUSUM being zero is very different however, as the CTCUSUM has a "hard" cut-off at zero, whereas the CTGLR has a "soft" cut-off at zero, in the sense that the MLE $\hat{\theta}_t$ can build up a buffer and thus the "true" value of the CTGLR can drop below zero.

- In Lemma 6.4.3 it was shown that a functional of the maximum likelihood estimate $\hat{\theta}_t$ is asymptotically normal. Heuristically this means that $\hat{\theta}_t$ converges asymptotically in mean towards the true value μ, with variance as specified in the statement. This means that as t becomes large, we are guaranteed that the estimate will be within a certain range of the true parameter with large probability.

- The main result of part I is the asymptotic distribution of the CTGLR chart stated in Theorem 6.5.1. This theorem tells us that a functional of the CTGLR chart converges asymptotically towards a normal distribution. Again, when t is large this can be used to determine an approximate range of values wherein $GLR(t)$ will lie with large probability. This is what we will use to determine an approximate average run length for the CTGLR chart.

- Corollary 6.6.1 lets us determine an approximate average run length for the CTGLR chart with control limit h, denoted by $ARL_{GLR}(\mu, h)$, when the true rate of failure μ is greater than zero. This theoretical ARL was later assessed in section 7.3 and found to be relatively accurate, even for small(er) values of t.

- In Lemma 6.8.1 asymptotic bounds (when all patients outcomes have been observed) for the expected value of the CTGLR are determined. Later, in section 7.4 we later find by means of simulation that the lower bound is extremely sharp, while the upper bound is extremely loose. In practice these bounds will only be useful for applications where the failure rate is extremely large, as their interpretation is only useful when all patients have failed.

- Corollary 6.10.2 tells us that the approximate asymptotic ($t \to \infty$) average run length of the CTGLR is always smaller than or equal to the asymptotic approximate average run length of the CTCUSUM, with equality when θ is chosen to be μ, with μ the true increased rate of failure at the hospital in consideration. In reality, as $\hat{\theta}_t$ needs time to converge towards the true value μ, the CTGLR will not have a shorter, or even the same run length as the CTCUSUM in the case that the value of θ is chosen correctly. When the disparity between θ and μ is large however, it

was shown by means of simulation in section 7.3 that the CTGLR enjoys a shorter ARL than the CTCUSUM.

11.4 CTMAXGLR

For the CTGLR chart we considered the hypotheses $\mu = 0$ against $\mu = \theta$ with $\theta > 0$ unknown, where the hospital has an increased rate of failure from the start of observations. We would now like to determine a chart which also doesn't require a pre-defined value of θ (as the CTGLR chart), and considers the hypotheses of a change of quality starting from some patient ν (as the (CT)CUSUM chart). This is very easily achieved by maximizing the CTGLR chart over the last $n-k$ observations, yielding the following definition.

Definition 11.4.1. *The Continuous Time Maximized Generalized Likelihood Ratio chart (CTMAXGLR) is given by:*

$$MAXGLR(t) = \max_k \hat{\theta}_{k,t} N_k^D(t) - (e^{\hat{\theta}_{k,t}} - 1) \sum_{i=k}^{n} \Lambda_i(t) \tag{87}$$

where the subscript k indicates that the quantities are taken over the last $n-k$ patients. This chart is used to test the hypothesis that the cumulative intensity at an institution differs by a factor of e^θ (with $\theta > 0$ unknown) from the null intensity $\Lambda_i(t)$, starting from some unknown patient ν:

$$\begin{aligned} H_0 :& X_1, X_2, ... \sim \Lambda_i \\ H_1 :& X_1, ..., X_{\nu-1} \sim \Lambda_i \\ & X_\nu, X_{\nu+1}, ... \sim \Lambda_i^\theta \end{aligned}$$

The counting processed are as defined in section 5.1. The maximum likelihood estimator $\hat{\theta}_k$ (limited to the last $n-k$ patients) was found in Lemma 6.2.1:

$$\hat{\theta}_{k,t} = \max\left(0, \ln\left(\frac{N_k^D(t)}{\sum_{i \geq k} \Lambda_i(t)}\right)\right). \tag{88}$$

*The null hypothesis is rejected at time t when $MAXGLR(t) \geq h$ for some $h > 0$, called the **control limit**.*

The greatest drawback of the CTMAXGLR chart is that it is very computationally intensive, as it not only requires us to recalculate the value of the chart at every time-point, but it also requires us to maximize over the last $n-k$ patients. Suppose we have n patients at time t. Then to determine the value of the CTMAXGLR chart we have to compute the value of n CTGLR charts. As we are interested in determining the chart at every time, the computational requirements grow as the chart is constructed over longer and longer time frames.

11.4.1 Properties

We state some properties we derived of the CTMAXGLR chart, most of which are generalizations of properties earlier found for the CTGLR chart (see also section 11.3.3).

- Similarly to the CTGLR chart, the CTMAXGLR chart is also non-negative. This follows easily from the fact that the CTMAXGLR is a CTGLR chart, taking into consideration only the last $n-k$ observations. Similarly, the strictly positive property holds only if $N_k^D(t) > \sum_{i \geq k} \Lambda_i(t)$ where k is such that the CTMAXGLR is maximized over the last $n-k$ patients.

- Even though we did not find an approximate average run length for the CTMAXGLR, Proposition 6.9.1 tells us that the average run length of the CTMAXGLR chart is bounded from above by the average run length of a CTGLR chart starting from patient ν plus τ, assuming that observations are out of control starting from patient ν and τ is the time of entry into the study of patient ν. This means that the approximate average run length found for the CTGLR in Corollary 6.6.1 lets us determine an approximate upper bound on the average run length of the CTMAXGLR, also when observations are not out of control from the beginning.

- Similarly to before, Corollary 6.10.2 now tells us that asymptotically the approximate ARL of the CTMAXGLR is smaller or equal to the approximate ARL of the CTCUSUM, even when the process is not initially out of control. This means that the CTMAXGLR is a suitable alternative for the CTCUSUM, in the case that the true value μ is not known.

Having introduced three continuous time inspection schemes, we are now interested in applying these charts on a real-life data set, which will be done in the next section. An example of all three charts introduced in this section can be found in Figure 7.

The key message of this section is that testing for a change in the quality of transplantations (in a real-life scenario) should be done either by the CTCUSUM chart or the CTMAXGLR chart, where the CTCUSUM is preferred when the true value θ of the alternative hypothesis is known, and the CTMAXGLR when it is not known or when it is variable over time. The CTCUSUM is very easy (computationally) to construct, while the CTMAXGLR is very hard to construct. Besides this, the CTMAXGLR requires time before its parameter estimates converge towards the true value. Due to the computational difficulties with the CTMAXGLR, we consider the CTGLR which is easier to construct and provides us with upper bounds on the run length of the CTMAXGLR.

12 LROI data set

The practical goal of this article is to apply the methods discussed and developed in this thesis on a data set from the Dutch Arthroplasty Register (LROI) [1], which will be done in this section. The LROI collects data about all orthopaedic implants performed in hospitals located in the Netherlands and provides this data to research groups in the hope of improving the quality of orthopaedic care. Due to the sensitive nature of the data provided by the LROI the data set used in this thesis is not publicly available, but can be requested for research purposes by applying to the LROI directly.

12.1 Description

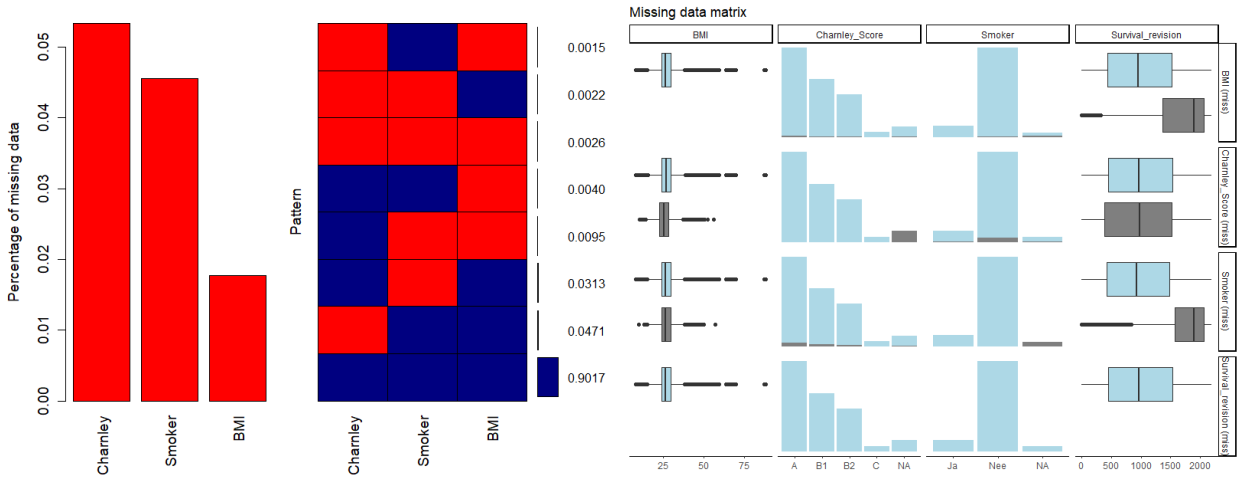
The data set [1] we will be using consists of information on total hip replacement surgeries. It contains information about 182385 patients over 97 different anonymized institutions (hospitals) across the Netherlands, ranging from large public hospitals to small private clinics, with the amount of relevant patients ranging from 800 to 1 per year. The hospitals are ordered according to the amount of surgeries taking place, with hospital 1 the "largest" and hospital 97 the "smallest". The data set contains information about all primary procedures at said hospitals occurring between 01/01/2014 and 01/01/2020 and follow up (revision, death or censoring) until 01/01/2020. For every patient we know the date of the primary procedure (when the replacement surgery took place) and the time (in days) until either a revision of the implant took place, the individual experienced death or the observation was censored. Additionally, we also know the time until death after a revision if this happened before the observation was censored. Besides this the data set contains some patient characteristics which are summarised in Table 4. Notable is that in six years only 2.4 percent of the subjects experienced a revision, while 1.7 percent of the subjects experienced a revision in their first year post surgery, indicating that the rate of revisions is extremely low, and most revisions happen in the first year post transplant.

12.2 Missing data

Only three variables had more than 0.5% of their values missing, which were BMI (1.8%), Smoking indicator (4.5%) and Charnley Score (5.3%), as can be seen in figure 14a. For the Charnley Score we know that a lot of patients could not be classified into either of the 4 categories. None of the patients which have had a revision take place during the 6 year period had any missing values. After visually inspecting the missing pairs plots for all variables combined with the survival times for patients with missing characteristics in Figure 14b we noticed that BMI and Smoking indicator were missing almost exclusively for patients who have a high survival time (meaning the individuals had their primary revisions in the first few years). This is likely due to the fact that hospitals only started recording BMI and smoking status in recent years, and that this was not done consistently in the beginning. Using the R package mice [35] we imputed missing values using predictive mean matching for age and BMI, logistic regression for Gender and Smoking indicator and polytomous regression for ASA classification and Charnley Score.

12.3 Risk-adjustment

Normally risk-adjustment models (such as logistic regression and cox proportional hazards models) are determined using some data set which is known to be in control. As we do not know which hospitals were up to standard in their treatment and at what times this occurred, it is not possible for us to determine a completely desirable model. We surmise however that due to the longitudinal nature of our data it is likely that there will be a balance between in control, out of control and above standard



(a) Pattern (b) Pairs

Figure 14: (a) Percentages of missing data for the 3 variables which have more than 0.5 % of their values missing (left) and the patterns (combinations and their proportions) of missing data (right). (b) Missing data pairs and the survival distribution of the individuals with missing data (in bold). The columns indicate which variable is missing.

data, which in turn will lead the risk-adjustment coefficients to balance out towards reasonable values. Another way to look at it is to consider the risk-adjustment models as an indicator of a nation-wide average instead of an in control situation. The above statement then comes down to believing that the national average is up to the desired standard.

In their article, van Schie et al. [11] determine a logistic regression model using three years of data which they use to find the outliers in a funnel plot at 3 years and a monthly Bernoulli CUSUM plot, both using revision 1 year post transplant as binary outcome. We will use a combination of data from the first 3 years and all 6 available years to construct risk-adjustment models, and compare the results in the coming sections.

After constructing a logistic regression model on the whole data set using 1−year post transplant revision as outcome and using all covariates in Table 4 we performed some analyses on whether the assumptions of the logistic regression model were met. First of all, using Wald tests we found that all the covariates had a significant effect on the model, even after applying a Bonferonni correction. Using the R package car [36] we calculated the variance inflating factor for all covariates and found no indication for the presence of multicollinearity. The linearity assumption was also assessed for covariate age by plotting a linear least squares regression equation against the logit of the predicted values. We found that this assumption was not met, as the slope differed depending on the predicted value. It is possible to adjust for this by creating splines over the age variable, but this makes the coefficients difficult to interpret. Besides this, as in the intended use of the methods the risk-adjustment model will be constructed multiple times over different time spans, this method of “fixing” the assumptions will not work in a general application.

Similarly, we constructed a Cox proportional hazards model over the whole data set, including all available covariates (see Table 4). Performing Wald tests for the determined coefficients we found that all variables except for age were significant predictors of survival, even after applying a Bonferonni correction. Using R package survminer [37], we test the proportional hazards assumption using the function cox.zph to find that the proportionality assumption does not hold for age, BMI, ASA Classification and Charnley Score. Graphically inspecting a plot of scaled Schoenfeld residuals against

N = 97 hospitals	Median (IQR)	Range
Continuous Variables		
Mean age (years)	69.5 (66.8 - 70.2)	51.8 - 81.5
Mean BMI (kg/m^2)	27.2 (26.9 - 27.5)	21.2 - 28.4
Discrete Variables		
Gender (%)		
Female	65.3 (63.2 - 67.1)	17.5 - 100
Male	34.7 (32.9 - 36.8)	0 - 82.5
Smoking (%)		
Yes	11.6 (10 - 13.4)	0 - 18.4
No	88.4 (86.6 - 90)	81.6 - 100
ASA Classification (%)		
I	15.5 (13.1 - 20.4)	0 - 53
II	63.7 (59.4 - 68.2)	43.8 - 93.8
III-IV	19.2 (12.6 - 24.6)	0 - 50
Charnley Score (%)		
A	47.1 (41.3 - 51.1)	0 - 76.7
B1	29 (25 - 33)	7.1 - 50
B2	21.5 (19.4 - 23.4)	5.5 - 50
C	2.3 (1.2 - 3.8)	0 - 16
Diagnosis (%)		
Osteoarthritis	86.9 (83.7 - 90.3)	0 - 98.8
Not Osteoarthritis	13.1 (9.7 - 16.3)	1.2 - 100
Statistics		
Procedures (number)		
in 3 years	756 (454 - 1227)	0 - 2523
in 6 years	1638 (1036 - 2462)	2 - 5093
Revision (%)		
1 year	1.7 (1.1 - 2.3)	0 - 10.4
end of follow-up	2.4 (1.6 - 3.3)	0 - 13.2

Table 4: Description of the data set

transformed time (using the function ggcoxzph from the survminer package, see Figure 19) we find that there is no noticeable pattern with time for these covariates. We argue that with a data set as large as ours, there will always be significant indications for a deviation from the ideal. It is possible to correct for this using f.e. time-varying coefficients, which can be done using the R package timereg [38]. Using the same reasoning as above, doing so would make the variables hard to interpret, as well as causing inconsistencies when constructing a model multiple times.

In conclusion, we argue that even though the statistical tests show that some assumptions are violated, these violations are not significant enough to impede functioning of said methods.

12.4 Bernoulli CUSUM vs Funnel Plot

Van Schie et al. [11] have shown that earlier detection of outlier hospitals in the LROI data set [1] is possible using a Bernoulli CUSUM chart (see section 10) instead of a yearly funnel plot (see section 9) while keeping a good sensitivity and specificity. In their article they thus use the same data set as the one we have (there are some slight changes due to hospital mergers), only restricted to the first

three years (follow-up until 01/01/2016). We will assess their conclusions using the extra information available to us. To do this we construct a yearly funnel plot and Bernoulli CUSUM charts with $e^\theta = 2$ and $h = 3.5$ and $h = 5$ respectively starting from 01/01/2016 (until 01/01/2020). Both methods use 1 year post transplant revision as binary outcome. The logistic regression risk-adjustment model (used by both the funnel plot as well as the Bernoulli CUSUM charts) was determined 4 times, on the first of January of $2017, 2018, 2019$ and 2020 using all the available data at that time point. This means that we do not use an in control data set for the risk-adjustment models, the consequences of this will be discussed in section 12.4.1. Consequentially, the yearly funnel plot uses the most recent risk-adjustment model. The Bernoulli CUSUM charts over the year 2016 use the risk-adjustment model from the first of January 2017. This way we can compare the methods with a similar risk-adjustment procedure. As the funnel plot is not a sequential testing procedure, its detection times are limited to the moment at which it is constructed. The Bernoulli CUSUM chart can produce a signal at any time of outcome, which in the case of this data set is any day of the year. We round the detection times of the Bernoulli CUSUM chart to the nearest month, leaving us with Table 5. Additionally, we summarise the gain in detection speed in Table 6, considering all hospitals detected by the funnel plot in the first three years (Table 6a), and all hospitals detected by any chart in all six years (Table 6b).

We can see that detection of relevant hospitals occurs faster using the Bernoulli CUSUM chart with $h = 3.5$ than using the funnel plot, and both methods detect exactly the same hospitals. We can see that this is not true for the Bernoulli CUSUM chart with $h = 5$, as for example hospital 23 would be detected later than with the funnel plot. Van Schie et al. [11] prefer the $h = 5$ control limit over the $h = 3.5$ limit, as with $h = 5$ they found a better specificity with respect to the detections of the funnel plot. This is not the case in our table as we do not construct monthly risk-adjustment models for the Bernoulli CUSUM charts. We chose not to do this because this method per construction causes the Bernoulli CUSUM charts to have a lower specificity due to the fact that all observations are used to construct the risk-adjustment model, even the ones which were due to a decrease in quality at some hospital. This means that if many failures with some risk-factor were to happen f.e. in December of 2016, then the Bernoulli CUSUM chart constructed in October of 2016 would be way more likely to signal than the funnel plot in January of 2017 due to failures with this specific risk factor. We find that the Bernoulli CUSUM chart with $h = 3.5$ detects exactly the same 13 hospitals at the three year mark as the funnel plot, with all hospitals detected earlier, with a faster median detection rate of 9 month (IQR $6 - 12$ months), while the CUSUM chart with $h = 5$ only detects 11 of the 13 hospitals signalled by the funnel plot at the three year mark, and signals one of the missing hospitals post the three year mark.

In Table 5 some cells are coloured red. This means that a detection of this hospital did not take place with said method. We would like to determine why these discrepancies take place. For this we construct Figure 15:

- In Figure 15a we can see a Bernoulli CUSUM chart for Hospital 6, which was detected by both CUSUM procedures but not by the funnel plot. This hospital is a prime example of the way in which a hospital can build up a buffer against detection by the funnel plot. We can see that this hospital performed admirably in the first few years, until around the 1500 day mark a lot of revisions suddenly started taking place. As the hospital was likely performing above standard before this, its O/E ratio at that point was way lower than the national average. This allows for the hospital to have sudden (short) problems in its quality before the O/E ratio will rise significantly above the national average, leading either to no detection or a significant delay in detection. To restrict this from happening, a funnel plot considering only the last K years of data could be considered.

- In Figure 15b we can see a hospital which was detected by the funnel plot, but neither of the

CUSUM procedures found enough evidence to report it. Comparing this figure with Figure 15a we can see that this hospital has had quite some revisions over its timespan. Because of this, its O/E ratio was most likely always on the high end. We can see that around the 1200 day mark a lot of revisions piled up consecutively. This was likely what caused the O/E ratio of this hospital to surpass the acceptable boundaries. The Bernoulli CUSUM chart however interprets this in another way. We can see that the chart was at 0 before this rise in revisions and the chart does not pass the 3.5 mark, which means that we could also attribute this sudden rise to chance or a very short spike in quality reduction.

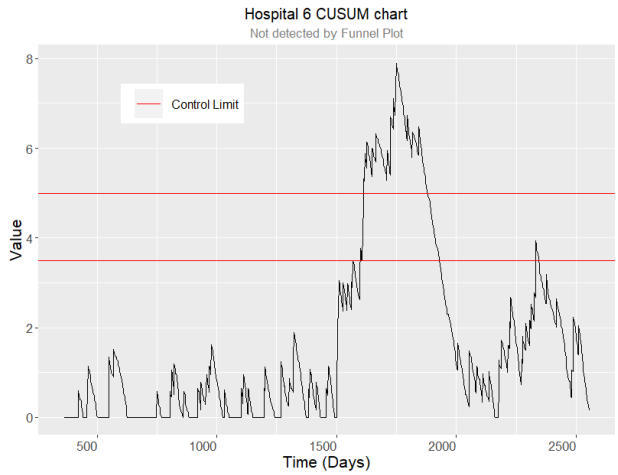
- In Figure 15c we can see Hospital 19, which was detected both by the funnel plot and the Bernoulli CUSUM with $h = 3.5$ boundary. We can see that the hospital was not doing well from the start, enough so for the Bernoulli CUSUM to warrant a signal at the $h = 3.5$ mark but not at the $h = 5$ mark. The funnel plot however is very sensitive to a lot of consecutive revisions, especially when we have only observed few outcomes, as shall be discussed further in section 12.4.1. We can see that the funnel plot detects this hospital at the three year mark (≈ 1095 days), which is the earliest detection time possible for the funnel plot using our method.
- Finally in Figure 15d we can see the Bernoulli CUSUM chart for Hospital 26, which was only detected by the $h = 3.5$ control limit. Again we can see that the hospital performed admirably in the first three years. Afterwards there seemed to be a prolonged but mild problem with the quality of care, which was not detected by the $h = 5$ control limit and not detected by the funnel plot due to the buffer built up in the first few years.

12.4.1 Discussion

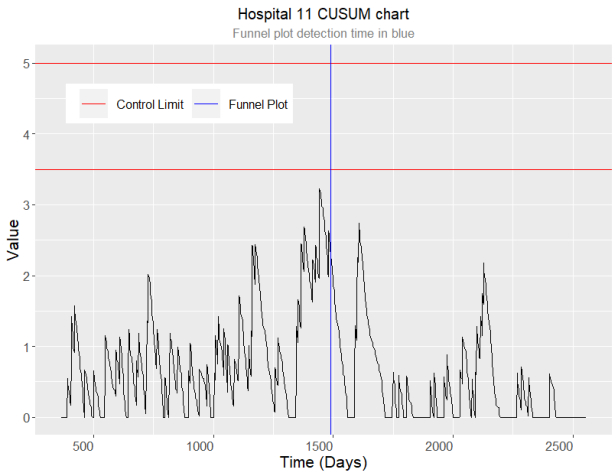
A very important distinction should be made between funnel plots and (CT)CUSUM charts. Whereas funnel plots are used to detect whether the O/E ratio of a hospital is significantly different from some national average (or in control ratio) at some point in time, a CUSUM procedure is used to sequentially test the hypothesis of a decrease in quality over time. Because of this, the funnel plot is not suitable as a continuous inspection scheme due to the immense risk of a type I error incurred by a multiple testing procedure. The choice to construct a funnel plot (bi-)yearly is therefore a decision which should be based on some sort of required detection speed, as well as a desired confidence in the possible detections. Another crucial matter to take into account is that the confidence bounds in a funnel plot depend on the amount of outcomes considered, which means that constructing a funnel plot including all hospitals at a certain time will not give all the hospitals sufficient time to converge towards a stable O/E ratio, especially when the rate of revision is very low. An example would be the smallest hospital in the data set, which has performed only two surgeries in six years. Suppose one of the two surgeries resulted in a revision, then retrieving the national average one year revision rate (1.7 percent) from Table 4, we would obtain (without risk-adjustment) an O/E ratio of $\frac{1}{2 \cdot 0.017} \approx 30$, which is well outside of both the 95 and 99 percent confidence intervals at 2 observations. Such "unlucky" occurrences would therefore result in detection by the funnel plot. This also means that a detection by the funnel plot is only really informative when we have observed a sufficient amount of outcomes. The CUSUM procedures however incorporate the situation described above by only signalling when a sufficient amount of failures are paired with an insufficient amount of favourable outcomes. An important quantity to adjust for this situation is the control limit h, which should be chosen in correspondence with the required detection speed and tolerable false detection rate. In light of this problem, the ARL (both in and out of control) of the (CT)CUSUM chart was shown in sections 10 and 11.2.1 respectively to depend on the amount of revisions per time unit, therefore warranting a different choice of the control limit h for hospitals with differing amount of patients.

Hospital nr.	Funnel plot $p = 0.95$ yearly	Bernoulli CUSUM $h = 3.5, e^\theta = 2$ monthly	Bernoulli CUSUM $h = 5, e^\theta = 2$ monthly	CTCUSUM $h = 9, e^\theta = 2$ monthly	CTMAXGLR $h = 15.1$ monthly
5	36	31	35	29	18
9	36	30	34	23	17
13	36	15	18	13	5
17	36	24	26	15	11
22	36	16	18	9	6
23	36	29	42	25	25
32	36	23	33	24	10
37	36	27	30	21	15
46	36	30	29	26	16
48	36	25	28	19	18
74	36	32	32	24	19
80	36	27	29	22	18
19	36	31		20	20
39	48	40	43	37	34
11	48			22	33
42	48			33	59
58	48			40	44
87	48				
81	60	52	64	58	
63	60	58		50	54
2	72	56	56	27	39
8	72	63	64	36	39
73	72				
41		43		56	32
26		44		54	38
29		48		37	20
6		52	52	37	39
18		60	61	43	45
55		61	69		50
50		62			48
4		64		54	47
60		66		66	50
35		70		43	52
31				50	66
7				57	44
40				58	
68				58	
77				66	
16					39
20					45
24					33
44					44
83					38

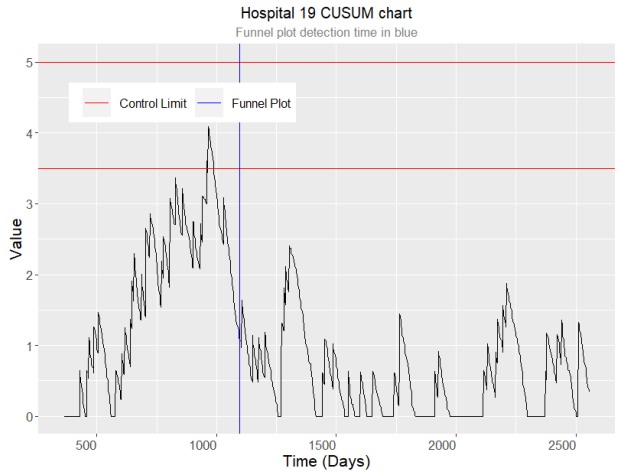
Table 5: Detection speed of charts in months on the LROI data set [1]. Red cells indicate that this method did not yield a detection on the corresponding hospital before 01/01/2020.



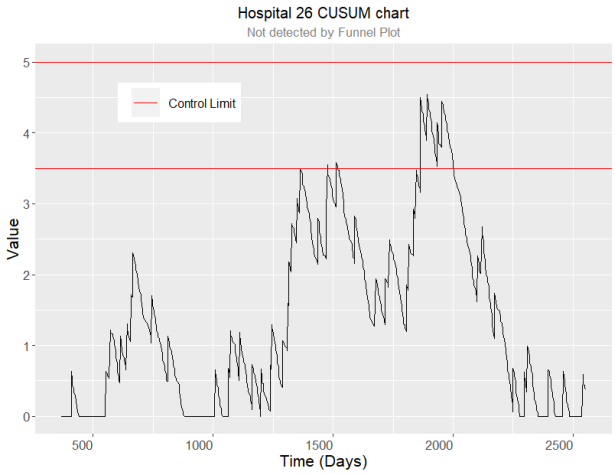
(a) Hospital 6



(b) Hospital 11



(c) Hospital 19



(d) Hospital 26

Figure 15: Bernoulli CUSUM charts for four hospitals with both the $h = 3.5$ and $h = 5$ control limits added in red. Additionally, a vertical blue line indicates the time at which the funnel plot would produce a signal for this hospital.

Median (IQR) difference in detection speed (months) over first 3 years

	Funnel plot	Bernoulli CUSUM	CTCUSUM	CTMAXGLR
Funnel plot	0 (0 - 0)	9 (6 - 12)	14 (12 - 17)	19 (18 - 25)
Bernoulli CUSUM	-9 (-12 - -6)	0 (0 - 0)	6 (4 - 7)	12 (10 - 13)
CTCUSUM	-14 (-17 - -12)	-6 (-7 - -4)	0 (0 - 0)	5 (3 - 8)
CTMAXGLR	-19 (-25 - -18)	-12 (-13 - -10)	-5 (-8 - -3)	0 (0 - 0)

(a) Hospitals detected in first three years by Funnel plot

Median (IQR) difference in detection speed (months) over 6 years

	Funnel plot	Bernoulli CUSUM	CTCUSUM	CTMAXGLR
Funnel plot	0 (0 - 0)	8 (6 - 12)	14 (11 - 21)	18 (15 - 25)
Bernoulli CUSUM	-8 (-12 - -6)	0 (0 - 0)	6 (2 - 11)	13 (10 - 14)
CTCUSUM	-14 (-21 - -11)	-6 (-11 - -2)	0 (0 - 0)	4 (-3 - 10)
CTMAXGLR	-18 (-25 - -15)	-13 (-14 - -10)	-4 (-10 - 3)	0 (0 - 0)

(b) All hospitals detected by any chart

Table 6: Difference in detection speed (months) of columns with respect to rows. Positive indicating quicker detection and negative indicating slower detection speeds.

Besides this, a funnel plot only indicates whether the O/E ratio at one specific time point is outside of some desired confidence interval. Due to this construction, the funnel plot can build up a buffer when failures happen at a smaller than national rate for an extended amount of time. To compensate for this, it is possible to consider only the last K years of data. This, in contrast, will favour some hospitals which had a bad performance in earlier time periods, as well as giving less confidence in the detection power of the funnel plot due to the lack of observations. Some of the pros and cons of the methods discussed in this section are summed in Table 7.

12.5 CTCUSUM & CTMAXGLR

Having compared the detection speed of the Bernoulli CUSUM chart and the funnel plot in section 12.4, we would now like to determine whether an additional improvement is possible using the Continuous Time CUSUM chart as introduced in section 5 or the CTMAXGLR chart as introduced in section 6.9. The CTCUSUM chart will be considered with $C = \infty$, meaning all revisions post primary procedure will be considered as qualifying, the consequences of this will be discussed in section 13.4.

12.5.1 CTCUSUM

Similarly to section 12.4 we construct the CTCUSUM chart (with $e^\theta = 2$) over the first three years of data using a Cox proportional hazards model fit on the data set restricted to all information up to the end of the third year (up until 01/01/2017). Then for the fourth year of data we use a Cox PH model fitted on the data set up until the end of the fourth year and so on. To determine a suitable value for the control limit h of the chart, we determine the smallest possible value of h such that all 13 hospitals which were detected by both the Bernoulli CUSUM as well as the funnel plot in the first three years of data are detected by the CTCUSUM chart as well, combined with the smallest possible amount of "false" detections. This resulted in a control limit of $h = 9$, with the CTCUSUM detecting all 13 required hospitals as well as 4 false positives. All false positives are however hospitals which were also detected by the funnel plot or Bernoulli CUSUM chart at a later time-point.

Using the $h = 9$ control limit we determine the detection times of the CTCUSUM chart in months since the start of the study, which can be found in Table 5. We can see that the CTCUSUM chart

outperforms the funnel plot and both Bernoulli CUSUM charts with respect to the detection time, except for hospital 32 which is signalled earlier by the Bernoulli CUSUM chart. In the first three years the CTCUSUM has a median faster detection rate of 6 month (IQR $4-7$) compared to the Bernoulli CUSUM and a median faster detection rate of 14 months (IQR $12-17$) with respect to the funnel plot (Table 6).

12.5.2 CTMAXGLR

We construct the CTMAXGLR using the same method as in section 12.5.1, yielding a control limit of $h = 15.1$. Notably the control limit for the CTMAXGLR is much larger than for the CTCUSUM ($h = 9$). This is because the CTMAXGLR determines an appropriate value for θ at every timepoint using the maximum likelihood estimate. Because the failure rate in this data set is so small the ML estimate $\hat{\theta}_t$ is very sensitive to consecutive failures, causing it to suddenly become very large when they happen, which in turn leads to a rapid rise in the value of the chart. To compensate for this we are therefore forced to choose a bigger control limit. Similarly to the CTCUSUM chart, the CTMAXGLR signals 4 "false positives" while signalling all 13 of the desired charts. These 4 hospitals are not the same for both charts, as the CTCUSUM "falsely" detects hospitals $2, 8, 11$ and 42, while the CTMAXGLR "falsely" detects hospitals $11, 29, 39$ and 41. All the false detections by the CTMAXGLR were also later on detected by either the funnel plot, the CTCUSUM or both.

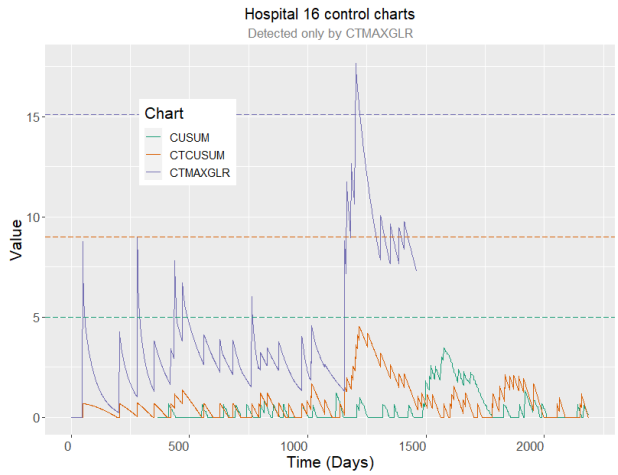
Using this $h = 15.1$ control limit we determine the detection times of the CTMAXGLR chart in months since the start of the study, which can be found in Table 5. The CTMAXGLR outperforms all charts in term of detection speed.

12.5.3 Visual examples

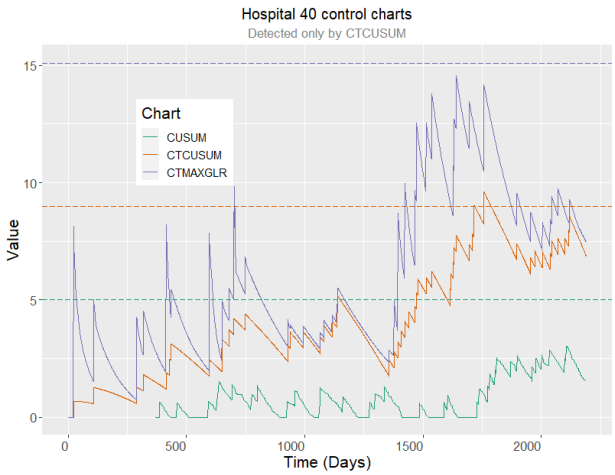
We would like to determine why some hospitals are signalled by some of the charts and not by others. For this we plot together the Bernoulli and continuous time CUSUM charts with the CTMAXGLR chart. The result can be seen in Figure 16.

- In Figure 16a a hospital which was detected only by the CTMAXGLR is plotted. Both the Bernoulli as well as the continuous time CUSUM charts also experience a spike in value (with the Bernoulli CUSUM having a delay of at most 1 year), but not as extreme as the CTMAXGLR chart. This is because the CUSUM procedures have a fixed value of θ, meaning they can rise only by a pre-determined amount. The CTMAXGLR chart however determined using maximum likelihood that the failures which happened at the time of its drastic rise were extremely unlikely and adjusted the value of $\hat{\theta}_t$ accordingly. This is one of the desirable properties of the CTMAXGLR, as we are interested not only in detecting a doubling of the revision rate, but in any increase in failures. The CUSUM procedures have a delay in detection compared to the CTMAXGLR in this specific case. Besides this, when the drastic increase in revision rate doesn't last long enough, the CUSUM procedures will not signal, as can be seen here.

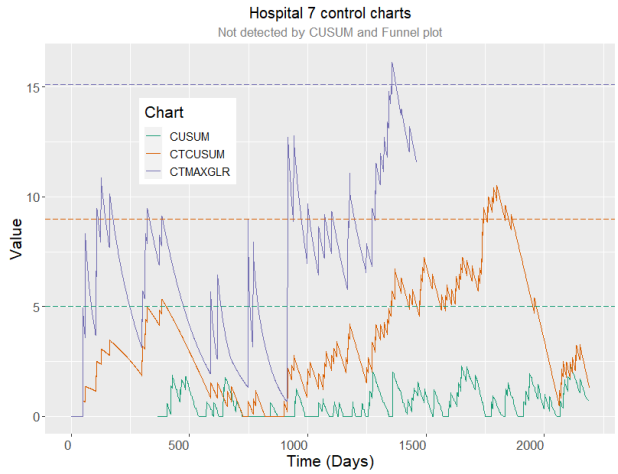
- In Figure 16b we can see the opposite of what happened in Figure 16a. The CTCUSUM chart is the only chart which produces a signal, while the CTMAXGLR chart is just a bit shy from its detection value. This is likely because the CTMAXGLR chart detected that the revision rate did not increase enough to warrant a signal. The CTCUSUM however is fixed at its $\theta = \ln(2)$ value and therefore keeps steadily increasing, even when the rate of failure might not have doubled. It is likely that the Bernoulli CUSUM chart would also have signalled, but due to the one year post-transplant outcome it has not yet processed the revisions which led to the signal in the CTCUSUM.



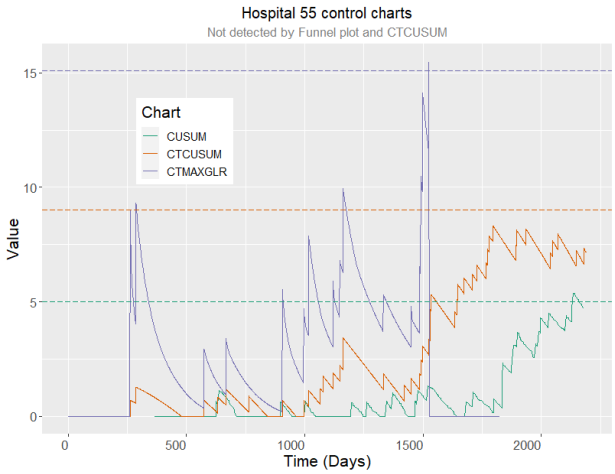
(a) Hospital 16



(b) Hospital 40



(c) Hospital 7



(d) Hospital 55

Figure 16: The (Bernoulli) CUSUM, CTCUSUM and CTMAXGLR charts for four hospitals with their control limits (dashed lines). The control limits can be found in Table 5. The CTMAXGLR was only constructed up until the year of detection to spare computation time.

	Funnel plot	Bernoulli CUSUM	CTCUSUM	CTMAXGLR
Pros	• Easy to interpret • Easy to construct	• Easy to interpret • Easy to construct • Sequential testing • No buffer • Possible to reset	• Easy to interpret • Real-time sequential testing • No buffer • Possible to reset	• Automated parameters • Real-time sequential testing • No buffer • Possible to reset
Cons	• Buffer build-up • Non-sequential • Cannot reset chart • Not real-time	• Not real-time • Delay in detection • Many assumptions	• Hard to construct • Many parameters • Many assumptions	• Hard to construct • Harder to interpret • Computationally intensive

Table 7: Some pros and cons of the considered methods.

- In Figure 16c we can see a hospital which was detected by both continuous time charts, but not by the Bernoulli CUSUM. Notably the Bernoulli CUSUM chart does not seem to rise at all. This is most likely because many of the revisions happened just past one year post transplant, leading to both the Bernoulli CUSUM not taking these revisions into account. This is another important drawback of the Bernoulli CUSUM (as well as the funnel plot).
- In Figure 16d we can see a hospital which was not detected by the funnel plot as well as the CTCUSUM chart. We can see that both the Bernoulli CUSUM as well as the CTMAXGLR deemed it necessary to signal this hospital, albeit only by a slight margin. This is likely due to the difference in risk-adjustment methods between the discrete and continuous time charts. Remember that the discrete time charts use a logistic regression model, with outcomes limited to one year post transplant, while the continuous charts use a Cox proportional hazards model. This detection could therefore be a false positive from both charts or it could be, similarly to Figure 16a, that the CTMAXGLR determined that the failures at some point were extremely unlikely and therefore adjusted its value of $\hat{\theta}_t$ accordingly.

12.5.4 Discussion

Summarizing above findings we note that the continuous time charts have an advantage over the discrete time charts as they use real-time outcomes instead of a dichotomized outcome $1-$year post transplant. This vastly improves detection speed (as we no longer have to wait a year), as well as allowing the chart to consider revisions which happen past the one year margin. Due to this, it becomes harder to compare the charts. This is especially pronounced in the fact that it is possible to detect exactly the same hospitals the funnel plot detects using the Bernoulli CUSUM chart, while the continuous time variants will always have "false" detections if we want to detect all 13 outliers. Continuing on this problem, the risk-adjustment procedures are also different. It is very difficult to compare the way these models influence the charts, as their assumptions are very different. It should be noted however, that the Cox PH model is also a continuous time model, which uses all the information available up to time t, whereas the logistic regression model can only use the observations which are already one year post transplant. In contrast to this, both models are based on various assumptions which are not necessarily true for this specific data set, especially as the data is of a longitudinal nature (i.e. hospitals use new techniques/new devices over time).

Secondly, the control limits for the continuous time charts were determined by minimizing detection speed, as well as detecting all 13 required hospitals while minimizing the amount of false positives. The control limit for the Bernoulli CUSUM chart was however chosen in accordance to van Schie et

al. [11]. Had we selected a control limit for the Bernoulli CUSUM chart using the same method as for the continuous charts, we would have arrived at $h = 3.3$ instead of $h = 3.5$. Even though this would improve the detection speed of the Bernoulli CUSUM chart slightly, the main reason why the continuous charts lead to quicker detection times is because they can use real-time outcomes and not due to this slight difference in parameter choice. Besides this, choosing a lower control limit for the Bernoulli CUSUM would potentially lead to more false detections past the 3 year mark.

Finally, the discrete and continuous time CUSUM charts use a fixed value of θ, whereas the CTMAXGLR chart can adjust this value to the current situation. This makes the CTMAXGLR chart way more versatile, especially when quick detections of a rapid rise in revision rate are required. As all the charts were trained on the detections by the funnel plot at the three year mark, this disparity was not very pronounced in the results of the first three years. Afterwards, as we saw in Figure 16, this difference in methods has caused clear distinctions in which hospitals were detected and how quickly they were signalled.

12.6 Conclusion

Summarily, we observed that in terms of detection speed the funnel plot is the least ideal method, followed by the Bernoulli CUSUM and then followed by the CTCUSUM and CTMAXGLR. The funnel plot, (CT)CUSUM and CT(MAX)GLR test different hypotheses. Hence, the goal of the researcher should be considered before deciding which of the charts to construct. Besides this, it is of importance to note that in the research above it was not clear which of the hospitals were in or out of control and at what times. Due to this it is unclear whether the way in which the control limits above were chosen are appropriate for the required results. Besides this, due to the way the risk-adjustment models were constructed the interpretation of the charts is not in relation to some in control revision rate, but with respect to a national average. As already stated, the parameters are chosen under the assumption that the national average is up to standard. Finally, the way in which the parameters were chosen and how the detection speeds were determined differ from what was done by van Schie et al. [11]. They constructed the Bernoulli CUSUM charts monthly, with a risk-adjustment model determined monthly as well. We believe that this construction is not mathematically sound, as this means that the funnel plot and Bernoulli CUSUM charts are compared to different national average rates of failure. Heuristically this can be seen as comparing a Bernoulli CUSUM chart with incomplete information against a funnel plot with complete information. Especially as the results of the Bernoulli CUSUM chart are trained on the results of the funnel plot, this is bound to cause problems for the choice of CUSUM control limit. The same reasoning also works for the CTCUSUM and CTMAXGLR charts. For this reason we only constructed the risk-adjustment models yearly. One could argue that this is unfair towards the funnel plot as it is only constructed yearly, while the Bernoulli CUSUM can detect monthly (or even daily), but it is important to keep in mind that this is indeed the biggest drawback of the funnel plot, as it is not a sequential procedure. The main pros and cons of all methods are summarized in Table 7.

13 Simulations

In section 12 we have applied four charts to a data set from the Dutch Arthoplasty Register (LROI) [1] and compared them primarily with respect to detection speed. Our largest problem in that section was that we did not have any reliable information about the times that hospitals had problems in their quality of care, and which hospitals had these problems at all. Due to this we constructed all the other charts with the hospitals detected by the funnel plot in the first three years as indication for problems in quality. We argued that the funnel plot is not a suitable method for performing sequential testing, as well as showing that all hospitals detected by the funnel plot can be detected earlier using a (continuous time) CUSUM chart. Finally, we also considered the CTMAXGLR chart, which could yield even quicker detection times, but suffered from some unique construction problems in the case of this particular data, due to the extremely low failure rate.

As we could not conclude anything about the true type I and II error probabilities of the newly introduced charts, many questions were left unanswered in the previous section. In this section we will assess these problems by means of simulation. A summary of the general simulation procedure can be found in section 15.4.

13.1 Research questions

As the funnel plot is not an acceptable method for performing sequential testing, we no longer consider this chart in the coming sections. We will primarily be interested in the following four research questions.

1. How powerful are the considered charts when the control limit h is chosen such that the type I error rate is approximately α over a certain span of time T?
2. How does the power and detection speed of the charts depend on the arrival rate ψ?
3. How do the CTCUSUM charts perform compared to the CT(MAX)GLR charts when the true rate of failure at a hospital is varying?
4. What is a realistic sensitivity and specificity for a data set such as the one from the LROI [1]?

Begun et al. [8] have developed a CUSUM procedure similar to the one introduced by Biswas & Kalbfleisch [2], but restricted only to Weibull and Gompertz distributed failure times. Both articles chose to select the control limit h by restricting the simulated type I error to 0.1 in 8 years and 0.15 in 5 years respectively, motivated by current procedures at medical centres in the US [7]. Following their reasoning we arrive at research question one. An important distinction between hospitals is that they have a different number of yearly procedures, determined by the size or specialization of the hospital. We saw in sections 11.2.1 and 11.3.3 that the approximate average run length of the CTCUSUM as well as the CT(MAX)GLR charts depends on the rate of arrival ψ, therefore it is important to determine how powerful our conclusions are for different sizes of hospitals, giving rise to research question two. A rather troublesome assumption of the CTCUSUM charts is the assumption that the alternative hazard rate differs by a fixed term e^{θ} from the in control hazard rate. In reality no hospital will have a constant failure rate at any time, leading to research question three. Finally, as our knowledge about the problems in the quality of care in the LROI data set [1] are extremely limited, we would like to determine a "realistic" sensitivity and specificity for future procedures by means of simulation, yielding research question four.

13.2 Power under varying ψ and type I error restriction

In this section we will address research questions one and two (13.1) by means of simulation. First we divide the hospitals in our data set into four distinct groups, according to their estimated rate of arrival $\hat{\psi}$ (determined using equation (8)), this is demonstrated in Figure 17a. We calculate their average estimated arrival rate, indicated in the figure in red. The hospitals are thus grouped into 4 distinct categories with arrival rates $\psi \in \{0.2, 0.6, 1, 1.7\}$. We then generate 500 hospitals for each value of ψ, by first determining the amount of patients arriving at each hospital in 6 years time and then bootstrapping patient characteristic from the full initial data set. Survival outcomes are then generated (see section 15.1) using a risk-adjusted Cox proportional hazards (non-parametric) model, which was determined using the R package survival [17]. First we generate hospitals under the null-hypothesis, so with $e^{\mu} = 1$. Using these in control hospitals we determine values of the control limit h for all four groups, such that the simulated type I error probability α is 0.1 in 6 years over the 500 samples. We chose this value of α because Begun et al. [8] chose a value of $\alpha = 0.1$ for an 8−year timespan for the similar NJR data set in the UK. Using this procedure we determined control limits shown in Table 8. Notice how different the control limits are for varying rates of ψ. This indicates that for practical considerations, one cannot take a single value of h for all hospitals in a data set.

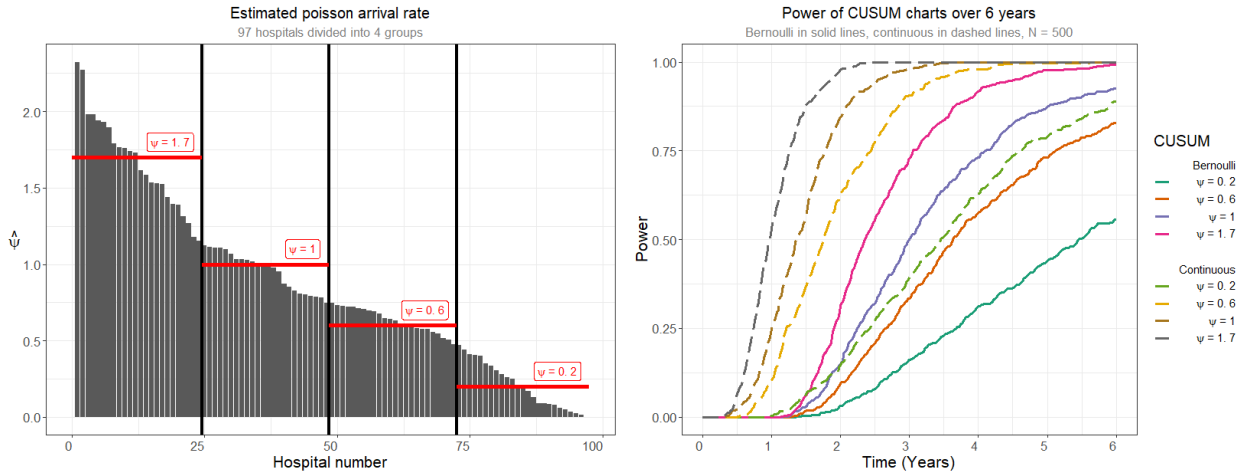
	Control limit h	
	CUSUM	CTCUSUM
ψ	$e^{\theta} = 2$	$e^{\theta} = 2$
0.2	2.19	2.62
0.6	3.4	5.59
1	3.88	7.17
1.7	4.53	9.11

Table 8: Control limits determined on a sample size of $N = 500$ in control ($e^{\mu} = 1$) observations such that the type I error in 6 years is equal to $\alpha = 0.1$. Both charts were constructed with $e^{\theta} = 2$.

Using a similar procedure we generated $N = 500$ out of control hospitals with $e^{\mu} = 2$ for the four values of ψ above. Then applying the Bernoulli CUSUM and CTCUSUM (with $e^{\theta} = 2$) to these hospitals with the found control limits we determined the simulated power of the charts at every time. The results can be found in Figure 17b. It is evident that the CTCUSUM is the more powerful method for all values of ψ considered. The angles of the lines are steeper for the CTCUSUM as well, indicating that the continuous chart needs less time to detect an out of control instance. Noticeably the Bernoulli CUSUM seems to have a delay of around 1 year with respect to the CTCUSUM. This is of course due to the outcome considered by the discrete time chart. All in all, the continuous time CUSUM is definitely the more powerful of the two charts, for all sizes of hospitals. Besides this, both charts are more powerful when the rate of arrivals is larger. Thus when considering a real-life data set, we should keep in mind that it is harder to detect out of control instances with a small arrival rate. This changes the interpretation of a signal, depending on the size of the hospital considered.

13.3 Varying rate of failure

In the previous sections, we always considered instances where $\mu > 0$ is some fixed value, mostly chosen such that $e^{\mu} = 2$ and assumed that we knew the true value so that we could chose $\theta = \mu$. In real life applications however, we cannot expect μ to be fixed, or to know the true value. For this reason, we performed a simulation study in part I to investigate the consequence of choosing θ wrongly. Similarly to above, in section 7.3 a value of h was determined such that the in control average run length was approximately 15 years, this time using a single value of ψ and a sample of $N = 3000$ (see Table 2).



(a) Estimated arrival rate (b) Power of (CT)CUSUM

Figure 17: (a) Estimated arrival rate as well as the subdivision of the hospitals into 4 groups. (b) Simulated power of the Bernoulli and continuous time CUSUM charts on a sample size of $N = 500$ out of control hospitals using control limit values such that the type I error $\alpha \approx 0.1$ in 6 years as in Table 8, with $e^{\mu} = 2$ and different arrival rates.

This was done for the CTCUSUM chart with $e^{\theta} = 1.4$ and $e^{\theta} = 1.8$, as well as for the CTGLR. As we saw in section 11.3.3, the CTGLR can build up a buffer under the null-rate, therefore the value of h was determined using the CTMAXGLR. As the CTMAXGLR is very computationally intensive, the sample size for determining h was restricted to $N = 120$. This means that the value of h for the CTGLR is less accurate, which can work both ways. Using the determined values of h, the ARL was calculated for $N = 3000$ out of control samples, with e^{μ} varying from 1 to 3. The results can be seen in Table 3.

We summarise the most notable results from section 7.3.

- It seems that the larger the true value of μ, the more merit there is in considering the CT(MAX)GLR instead of the CTCUSUM, especially when the value of θ for the CTCUSUM is chosen wrongly.
- The distribution of the ARL is right-skewed for all the charts, with the skewness being larger for the CTCUSUM charts. This indicates that there are more outliers using the CTCUSUM chart than the CT(MAX)GLR.
- The standard deviation of the ARL is smaller for the CTGLR chart than for the CTCUSUM charts when μ is small, while for large values of μ it is reversed. We hypothesise that this is due to the CUSUM charts having a fixed value of θ, whereas the GLR charts have a variable $\hat{\theta}_t$. We surmise that the CTMAXGLR might have a smaller standard deviation in its detection times, as it always has shorter detection times than a CTGLR chart.
- The theoretical values for the approximate average run length discussed in sections 11 seem to correspond with the observed average run lengths, even for relatively small values of t. The approximate average run length found by Biswas & Kalbfleisch [2] with $C = 1$ does not seem to correspond to the simulated values. This is because the ARL was determined from the start of observations instead of starting from $C = 1$ year post beginning of observations.

Keeping in mind above considerations, the CTMAXGLR should be the preferred chart over the CTCUSUM when the rate of failure is variable and even more so when the true value $e^{\mu} >> 1$.

13.4 Sensitivity and specificity analysis

In this section we would like to combine the first three considered research questions into a single simulation study where we try to create a data set which resembles a realistic scenario. For this we make the following assumptions.

- The true null cumulative hazard rate is Weibull distributed, with (risk-adjustment) parameters determined using the function phreg from the R package eha [39] on the full initial data set. A comparison between non-parametric (package survival [17]) and parametric hazard can be seen in Figure 18.
- Hospitals fall in 4 categories, grouped by the following values of ψ: $\psi \in [0.01, 0.46]$, $\psi \in [0.47, 0.74]$, $\psi \in [0.75, 1.12]$ and $\psi \in [1.13, 2.33]$. Each group corresponds to the way we divided the hospitals in our data set in Figure 17a. Each hospital in one of the groups has its value drawn from the indicated range uniformly.
- For every category of ψ above, we consider 300 hospitals where the proportion of hospitals which experience a reduction in the quality of care, indicated by p, is $0.2, 0.3$ or 0.4. For every group of ψ we therefore generate 1200 hospitals, with $K \sim Bin(300, p)$ the amount of hospitals which experience a reduction in quality in that group.
- Hospitals are only considered for 6 years after the start of study, and hospitals which were chosen to have an out of control instance have their first out of control patient drawn uniformly in the first 5.5 years.
- As the original data set contained censored observations and most of these censored observations were due to the death of individuals, we construct a proportional hazards failure model for death on the initial data set. Using this model we then determine times of death for each individual in the simulated data set independent of the time of revision, and censor the observation if death takes place before a revision.
- Whenever an observation is in control, its true value of failure e^{μ} is sampled from a $N(1, 0.01)$ distribution, as to introduce some variability in the in control data sets.
- Whenever an observation is out of control, its true value of failure e^{μ} is sampled form a $\mathcal{N}(2, 0.09)$ distribution.
- We take $e^{\theta} = 2$ for both the Bernoulli CUSUM as well as the CTCUSUM, so as to indicate that we have sufficient prior information about the expected rate of failure in the data set.

Begun et al. [8] have found that the competing risks of revision and death are independent in the NJR data set. As this data set is very similar in nature to ours, we expect the same to hold. Due to this, we can generate survival outcomes independent of revision outcomes to simulate a realistic censoring mechanism. In Figure 17a we divided the hospitals into four groups according to their arrival rate, but it is not realistic that all hospitals should have exactly the same arrival rate. To compensate for this we sample uniformly for each hospital in a certain range. Finally, we saw in Table 4 that around 40 percent of the hospitals were detected (considering detections by all charts). As we expect some of these to be false detections, we are interested in a failure proportion between 0.2 and 0.4 in the hospitals in question.

Having generated hospitals under above assumptions, we determine the sensitivity and specificity of the Bernoulli and continuous time CUSUM charts applied on these hospitals using the control limit values determined in Table 8. The results can be seen in Table 9. Due to the slightly lower sample

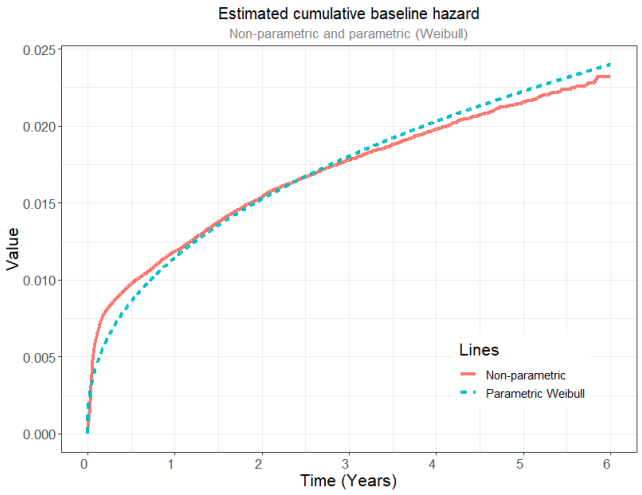


Figure 18: Estimated (parametric Weibull and non-parametric) cumulative baseline hazard for the full data set.

size we note that there is quite some variability in the values. Remember that the values of the control limits were determined so that the simulated specificity was around 0.9 for all the combinations. We immediately notice that the specificity of the Bernoulli CUSUM is very close to 0.9, while the specificity of the CTCUSUM is comparatively small. We thus conclude that the determined values of h for the CTCUSUM chart lead to a very good sensitivity, but not towards the required specificity causing many false alarms.

N = 300	Bernoulli CUSUM ($e^\theta = 2$)						
Daily		Sensitivity			Specificity		
arrival rate ψ	h	$p = 0.2$	$p = 0.3$	$p = 0.4$	$p = 0.2$	$p = 0.3$	$p = 0.4$
0.2	2.19	0.45	0.58	0.49	0.85	0.76	0.78
0.6	3.40	0.50	0.67	0.64	0.86	0.89	0.90
1	3.88	0.74	0.71	0.71	0.87	0.90	0.91
1.7	4.53	0.70	0.82	0.75	0.90	0.91	0.90
N = 300	Continuous Time CUSUM ($e^\theta = 2$)						
Daily		Sensitivity			Specificity		
arrival rate ψ	h	$p = 0.2$	$p = 0.3$	$p = 0.4$	$p = 0.2$	$p = 0.3$	$p = 0.4$
0.2	2.62	0.74	0.81	0.79	0.53	0.45	0.47
0.6	5.59	0.87	0.85	0.85	0.62	0.60	0.61
1	7.17	0.88	0.84	0.89	0.60	0.64	0.62
1.7	9.11	0.94	0.93	0.88	0.55	0.58	0.52

Table 9: Sensitivity and specificity of the Bernoulli and continuous time CUSUM charts for a "realistic" hospital data set, each value determined on $N = 300$ simulated hospitals. p indicates the probability of a hospital having an out of control period in that sample.

The most likely culprit for this very large drop in specificity in Table 9 for the CTCUSUM but not for the Bernoulli CUSUM is the censoring mechanism. There are two things at play here:

1. Only when an observation is censored less than one year post transplant will it have an influence on the Bernoulli CUSUM chart. For the CTCUSUM chart however, censoring of the observation

will have an influence on the chart at any time past primary procedure. Therefore censored observations always have less influence on the Bernoulli CUSUM chart compared to the continuous time variant, given that censoring can happen later than one year post transplant.

2. If an observation is censored less than one year post transplant, the Bernoulli CUSUM chart considers the outcome to be favourable (no revision). This means that censored observations make the Bernoulli CUSUM chart decrease in value. Censored observations in the CTCUSUM chart stop providing information to the chart, meaning that their downward drift (right side of equation 77) is eliminated from the chart starting from the time of censoring. This means that the downward slope of the chart decreases, resulting in a net increase in the value of the chart.

Above considerations explain why the Bernoulli CUSUM does not experience a significant decrease in specificity, while lacking sensitivity due to possibly unfavourable outcomes being considered as favourable. In contrast, it is clear that the censoring mechanism increases the net value of the CT-CUSUM by virtue of decreasing the downward slope, leading to a worse specificity, while not impairing the sensitivity. It is therefore very important to note that censoring mechanisms affect the Bernoulli and continuous time CUSUM charts differently.

To compensate for this, the practical value of h should be chosen higher for the CTCUSUM (to increase the specificity), and if desirable also for the Bernoulli CUSUM (but only slightly as the specificity was hardly imparted). We try to quantify how much higher this value should be for the CTCUSUM and the Bernoulli CUSUM by determining a value for h such that the specificity averaged over the three possible values of p is approximately 0.9. The result can be found in Table 10. Notice that we had to increase the value of h for both the charts, but proportionally the increase was much larger for the CTCUSUM. The most notable result from the table is that with an equal simulated specificity, the CTCUSUM seems to have a better sensitivity for low values of ψ, but their sensitivities do not seem to differ significantly for higher values of ψ. This is in line with the observations of Biswas & Kalbfleisch [2]. They found that the False Discovery Rate (FDR), which is a measure of the proportion of false discoveries, was only significantly different for low values of ψ as well.

In contrast to this, considering all out of control hospitals which were detected by both charts, the time to detection from the first out of control observation was on average shorter by 76 days for the CTCUSUM, which is about 2.5 months. We thus conclude that the CTCUSUM enjoys faster detection times for all hospitals, coupled with a better sensitivity for small hospitals.

In conclusion, the value of the control limit h for the Bernoulli CUSUM can be chosen with disregard to censored outcomes, given that most observations are censored later than the follow-up of the chart. For the CTCUSUM however it is important to include a suitable censoring mechanism when determining a value of h, as the charts specificity is heavily impacted by censoring.

N = 300	Bernoulli CUSUM ($e^\theta = 2$)						
Daily		Sensitivity			Specificity		
arrival rate ψ	h	$p = 0.2$	$p = 0.3$	$p = 0.4$	$p = 0.2$	$p = 0.3$	$p = 0.4$
0.2	2.76	0.35	0.43	0.37	0.93	0.91	0.86
0.6	3.52	0.50	0.67	0.62	0.89	0.91	0.90
1	3.96	0.72	0.71	0.70	0.88	0.91	0.91
1.7	4.49	0.70	0.82	0.75	0.90	0.90	0.90
N = 300	Continuous Time CUSUM ($e^\theta = 2$)						
Daily		Sensitivity			Specificity		
arrival rate ψ	h	$p = 0.2$	$p = 0.3$	$p = 0.4$	$p = 0.2$	$p = 0.3$	$p = 0.4$
0.2	5.25	0.40	0.54	0.47	0.94	0.87	0.89
0.6	8.37	0.75	0.72	0.71	0.91	0.90	0.89
1	10.89	0.75	0.71	0.70	0.89	0.90	0.91
1.7	15.27	0.76	0.75	0.73	0.88	0.93	0.89

Table 10: Sensitivity and specificity of the Bernoulli and continuous time CUSUM charts for a "realistic" hospital data set with control limits h adjusted such that the mean specificity over a value of ψ is 0.9.

14 Discussion and recommendations for practice

In this thesis we have applied the funnel plot, Bernoulli CUSUM, continuous time CUSUM and continuous time MAXGLR chart on a data set provided by the LROI [1] in section 12. We found that in the first three years it is possible to detect the same hospitals using the Bernoulli CUSUM chart with $h = 3.5$ control limit as with the currently employed funnel plot, which is currently employed. The Bernoulli CUSUM chart had a median 9 months smaller detection time for the hospitals in question. Consecutively, the CTCUSUM was able to improve on this detection time by another 6 months (median), with the CTMAXGLR improving detection speed even further by 5 months (median). Past the three year margin, the hospitals detected by the methods varied largely. From the hospitals detected by the funnel plot past the three year margin, five were not detected by the Bernoulli CUSUM, two were not detected by the CTCUSUM and three were not detected by the CTMAXGLR.

We argued that constructing yearly funnel plots is not a suitable inspection scheme for the problem at hand in section 12.4.1, mostly due to its non-sequential nature of testing. While the Bernoulli CUSUM chart is a suitable method, it is not well adjusted to survival outcomes present in the data set at hand. Consequently our preference goes out to the continuous time inspection schemes introduced in part I of this thesis: the CTCUSUM and CTMAXGLR. In section 13.2 these were found to have a smaller detection time for out of control instances, as well as being more powerful under a fixed type I error probability, especially for (small) hospitals with low arrival rates.

Extensions of the CTCUSUM and CTMAXGLR could be considered, for example by adding Frailty terms (compensating for unobserved variables) or competing risks to the model. Besides, procedures for constructing risk-adjustment models could be developed for a data set similar to ours, where it is unknown which hospitals have experienced problems in their quality of care.

To further reduce detection times, quick-start procedures such as the one proposed by Lucas & Crosier [30] could be considered. They found that for the discrete time CUSUM chart applied to normally distributed data, picking a head-start of $h/2$ yielded faster out of control detection times, while keeping the in control detection times approximately equal. These can then also be incorporated after a detection has taken place, as a manner of resetting the charts.

The data set available to us contained a lot of individuals which had their outcomes censored due to experiencing death. To interpret this in the chart, competing risks methodology could be added to the CTCUSUM and CTMAXGLR. Begun et al. [8] have found that competing risks of death and revision were independent in the English National Joint Registry data set. This suggests that death should not be included in the model, as only the quality of revision procedures is of interest. We believe however that this decision should be made in accordance with medical professionals, as they have more prior information about the connection between revision and death.

As this thesis was focussed on comparing Bernoulli CUSUM charts with continuous time CUSUM/GLR charts, we ignored other available continuous time charts for survival data. Future (simulation) research should be done to compare the CTCUSUM and CTMAXGLR with other continuous time monitoring schemes such as the RAST CUSUM by Sego et al. [31], uEWMA for survival time data by Steiner & Jones [32] and finally the STRAND chart by Grigg [33]. Grigg argues that the biggest drawback of the CTCUSUM is the absence of a shuffling mechanism for patients and absence of weighted observations. Because of this clusters of failures could be considered as noise leading to a delay in detections as well as ignoring the fact that more recent observations are of more interest to detection. Both these issues are addressed in the uEWMA and STRAND charts.

14.1 Recommendations for practice

Some practical recommendations for the Dutch Arthroplasty Register are made in this section. Whereas the focus in this thesis lies on arthroplasty outcomes, the recommendations made in this section are more generally applicable to most practical situations where survival outcomes are of interest. Firstly, we recommend to replace funnel plots by control charts where possible, as the former are not suitable for performing sequential tests for a reduction in the quality of care. Depending on the amount of expert knowledge about the expected change in the rate of failure, either the CTCUSUM or CTMAXGLR chart should be employed for future monitoring of hospitals. The CTCUSUM should be used when there is sufficient confidence in the knowledge about the true change in failure rate. The CTMAXGLR is a more general method not requiring such knowledge, but is harder to interpret for laymen and requires a lot of computational time to determine parameters.

Most importantly, regardless of which chart is employed, control limits should be chosen with respect to the size of the hospital. It is possible to achieve this by splitting hospitals into distinct groups according to the amount of patients per time unit, as was done in this thesis. The amount of groups should be chosen by an expert on hospital matters, who knows which hospitals are comparable. Employing any of the continuous time charts with a single control limit will lead to delays in detections for small hospitals and many false detections for large hospitals.

Values for control limits h should be determined according to the methods in this thesis. For the Bernoulli CUSUM, patient characteristics can be bootstrapped from the data set to determine a control limit such that the type I error over T years is approximately some desired value α, without including a censoring mechanism into the simulation. The value for the control limit was shown to work in a "realistic" application including right-censored observations. For the CTCUSUM, we saw that it is no longer possible to ignore censoring mechanisms when determining a value for the control limit h. It is therefore important to decide on a suitable censoring mechanism before performing a simulation study. An appropriate value for the control limit can then be found similarly to the Bernoulli CUSUM, by limiting the type I error over T years by α.

Finally, to further increase detection speed, risk-adjustment models should be constructed on hospital data which is known to be in control. We realise that many hospitals do not release this information due to its sensitive nature, but determining a reliable in control hazard rate is extremely likely to produce quicker detection results than a national average rate of failure.

References

[1] URL `https://www.lroi.nl/wetenschap/onderzoek-met-lroi`.

[2] P. Biswas and J. D. Kalbfleisch. A risk-adjusted CUSUM in continuous time based on the Cox model. *Statistics in Medicine*, 27(17):3452–3452, 2008. doi: 10.1002/sim.3296.

[3] E. S. Page. Continuous inspection schemes. *Biometrika*, 41(1/2):100, 1954. doi: 10.2307/2333009.

[4] D. Seland. A brief history of statistical process control, Jan 2021. URL `https://www.qualitymag.com/articles/96349-a-brief-history-of-statistical-process-control`.

[5] A. Wald. Sequential tests of statistical hypotheses. *The Annals of Mathematical Statistics*, 16(2): 117–186, 1945. doi: 10.1214/aoms/1177731118.

[6] D. Spiegelhalter. Risk-adjusted sequential probability ratio tests: applications to Bristol, Shipman and adult cardiac surgery. *International Journal for Quality in Health Care*, 15(1):7–13, 2003. doi: 10.1093/intqhc/15.1.7.

[7] D. M. Dickinson, T. H. Shearon, J. Okeefe, H. H. Wong, C. L. Berg, J. D. Rosendale, F. L. Delmonico, R. L. Webb, and R. A. Wolfe. SRTR center-specific reporting tools: Post-transplant outcomes. *American Journal of Transplantation*, 6(5p2):1198–1211, 2006. doi: 10.1111/j.1600-6143.2006.01275.x.

[8] A. Begun, E. Kulinskaya, and A. J. Macgregor. 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(1), 2019. doi: 10.1186/s12874-019-0853-2.

[9] C. Kok. How statistics can save lives. Master's thesis, Leiden University, the Netherlands, 2019.

[10] LROI. Magazine LROI 2018. URL `https://issuu.com/novlroi/docs/magazine_lroi_2018`.

[11] P. van Schie, L. Van Bodegom-Vos, L. Van Steenbergen, R. Nelissen, and P. Marang-van de Mheen. Monitoring hospital performance with statistical process control after total hip and knee arthroplasty: A study to determine how much earlier worsening performance can be detected. *Journal of Bone and Joint Surgery*, Publish Ahead of Print, 09 2020. doi: 10.2106/JBJS.20.00005.

[12] J. P. Klein and M. L. Moeschberger. *Survival analysis: techniques for censored and truncated data*. Springer, 2011.

[13] D. R. Cox. Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2):187–202, 1972. doi: 10.1111/j.2517-6161.1972.tb00899.x.

[14] A. A. Tsiatis. A large sample study of Cox's regression model. *The Annals of Statistics*, 9(1): 93–108, 1981. doi: 10.1214/aos/1176345335.

[15] N.E. Breslow. Discussion of the paper by D.R. Cox J R. *Journal of the Royal Statistical Society*, 1972.

[16] B. Efron. The efficiency of Cox's likelihood function for censored data. *Journal of the American Statistical Association*, 72(359):557–565, 1977. doi: 10.1080/01621459.1977.10480613.

[17] Terry M. T. and Patricia M. G. *Modeling Survival Data: Extending the Cox Model*. Springer, New York, 2000. ISBN 0-387-98784-3.

[18] O. Aalen. Nonparametric inference for a family of counting processes. *The Annals of Statistics*, 6(4):701–726, 1978. doi: 10.1214/aos/1176344247.

[19] L.C.M. Kallenberg. Stochastische besliskunde. 2020.

[20] J. Rice. *Mathematical statistics and data analysis*. W. Ross MacDonald School Resource Services Library, 2015.

[21] B. D'Auria. Lecture notes in stochastic processes, Feb 2012.

[22] G. Lorden. Procedures for reacting to a change in distribution. *The Annals of Mathematical Statistics*, 42(6):1897–1908, 1971. doi: 10.1214/aoms/1177693055.

[23] G. V. Moustakides. Optimal stopping times for detecting changes in distributions. *The Annals of Statistics*, 14(4):1379–1387, 1986. doi: 10.1214/aos/1176350164.

[24] D. Siegmund. *Sequential analysis*. Springer-Verlag, 1983.

[25] G. Lorden. Open-ended tests for Koopman-Darmois families. *The Annals of Statistics*, 1(4): 633–643, 1973. doi: 10.1214/aos/1176342459.

[26] D. R. Cox and H. D. Miller. *The theory of stochastic processes*. Chapman & Hall, 2009.

[27] B. Hoadley. Asymptotic properties of maximum likelihood estimators for the independent not identically distributed case. *The Annals of Mathematical Statistics*, 42(6):1977–1991, 1971. doi: 10.1214/aoms/1177693066.

[28] A. W. van der Vaart. *Asymptotic statistics*. Cambridge Univ. Press, 2007.

[29] G. Casella and R. L. Berger. *Statistical inference*. Brooks/COle Cengage Learning, 2017.

[30] J. M. Lucas and R. B. Crosier. Fast initial response for CUSUM quality-control schemes: Give your CUSUM a head start. *Technometrics*, 42(1):102–107, 2000. doi: 10.1080/00401706.2000.10485987.

[31] L. H. Sego, M. R. Reynolds, and W. H. Woodall. Risk-adjusted monitoring of survival times. *Statistics in Medicine*, 28(9):1386–1401, 2009. doi: 10.1002/sim.3546.

[32] S. H. Steiner and M. Jones. Risk-adjusted survival time monitoring with an updating exponentially weighted moving average (EWMA) control chart. *Statistics in Medicine*, 29(4):444–454, 2009. doi: 10.1002/sim.3788.

[33] O. A. Grigg. The STRAND chart: A survival time control chart. *Statistics in Medicine*, 38(9): 1651–1661, 2018. doi: 10.1002/sim.8065.

[34] D. J. Spiegelhalter. Funnel plots for comparing institutional performance. *Statistics in Medicine*, 24(8):1185–1202, 2005. doi: 10.1002/sim.1970.

[35] S. van Buuren and K. Groothuis-Oudshoorn. mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, 45(3):1–67, 2011. URL `https://www.jstatsoft.org/v45/i03/`.

[36] J. Fox and S. Weisberg. *An R Companion to Applied Regression*. Sage, Thousand Oaks CA, second edition, 2011. URL `http://socserv.socsci.mcmaster.ca/jfox/Books/Companion`.

[37] A. Kassambara, M. Kosinski, and P. Biecek. *survminer: Drawing Survival Curves using 'ggplot2'*, 2020. URL `https://CRAN.R-project.org/package=survminer`. R package version 0.4.8.

[38] T. H. Scheike and M. Zhang. Analyzing competing risk data using the R timereg package. *Journal of Statistical Software*, 38(2):1–15, 2011. URL `https://www.jstatsoft.org/v38/i02/`.

[39] G. Broström. *Event History Analysis with R*. Chapman and Hall/CRC, Boca Raton, 2012. doi: 10.1201/9781315373942.

[40] P. C. Austin. Generating survival times to simulate Cox proportional hazards models with time-varying covariates. *Statistics in Medicine*, 31(29):3946–3958, Apr 2012. doi: 10.1002/sim.5452.

[41] F Bijma, M. Jonker, and A. W. van der Vaart. *An Introduction to Mathematical Statistics*. Amsterdam University Press, 2018.

[42] H. Bauer. *Measure and Integration Theory*. De Gruyter, 2011.

15 Appendix

15.1 Cox proportional hazards

Austin et al. [40] have shown how to generate survival times under the Cox proportional hazards model. Suppose we have a distribution function for the survival time T, and the hazard rate under the Cox PH model given by:

$$h(t|Z) = h_0(t) \cdot e^{\beta Z}.$$

We then have for the distribution function that:

$$F(t|Z) = 1 - e^{-H_0(t)e^{\beta Z}}$$

therefore using that if $X \sim F$, then $F(X) \sim U[0,1]$ and $1 - U[0,1] \sim U[0,1]$, thus:

$$U = e^{-H_0(T)e^{\beta Z}} \sim U[0,1]$$

therefore recovering the survival time:

$$T = H_0^{-1}[-\ln(U)e^{-\beta Z}] \tag{89}$$

with $U \sim U[0,1]$. Further on in this thesis we will consider hypothesis tests where we want to check whether the hazard rate differs from the null rate by a factor of e^{μ} for some $\mu > 0$. In that case we have that $h^{\mu}(t|Z) = h_0(t)e^{\beta Z}e^{\mu}$ and correspondingly outcomes are generated using:

$$T = H_0^{-1}[-\ln(U)e^{-\beta Z}e^{-\mu}]. \tag{90}$$

Thus to generate survival outcomes from a known baseline (cumulative) hazard function we only need to be able to invert said function and generate uniformly distributed random variable outcomes. Luckily, generating uniformly distributed random outcomes can be done easily using many statistical software packages, thus we only require an invertible (cumulative) baseline hazard function.

15.2 Expectation over covariates

Lemma 15.2.1. *Let f_i^{μ} be a non-negative Borel-measurable function with primitive function F_i^{μ} and suppose that:*

$$\mathbb{E}_{Z_i}\left[f_i^{\mu}(u-x)\right]$$

exists. Then:

$$\int_0^u \mathbb{E}_{Z_i}\left[f_i^{\mu}(u-x)\right]dx = \mathbb{E}_{Z_i}\left[F_i^{\mu}(u)\right].$$

Proof. We can write:

$$\int_0^u \mathbb{E}_{Z_i}\left[f_i^{\mu}(u-x)\right]dx = \int_0^u \int_{\Omega_{Z_i}} f_i^{\mu}(u-x)dzdx.$$

As f_i^{μ} is non-negative and Borel measurable and the expected value exists (thus the double integral is finite) we obtain from Fubini's theorem in section 23 of Bauer [42] that we can switch the order of integration to obtain:

$$\begin{aligned}\int_0^u \mathbb{E}_{Z_i}\left[f_i^{\mu}(u-x)\right]dx &= \int_{\Omega_{Z_i}} \int_0^u f_i^{\mu}(u-x)dxdz \\ &= \int_{\Omega_{Z_i}} F_i^{\mu}(u)dz = \mathbb{E}_{Z_i}\left[F_i^{\mu}(u)\right]\end{aligned}$$

thus proving the statement. □

15.3 Expected value of A(t)

In this section we prove a result necessary for the approximation of the ARL of the CTCUSUM and CTGLR charts in sections 5.4 and 6 respectively. We show that the expected value of $A(t)$ is equal to $\int_0^t \gamma_u du$.

Lemma 15.3.1. *Let $A(t) = \sum_{i\geq 1} \Lambda_i(t)$ with h_i^μ and f_i^μ non-negative Borel measurable functions and $\mathbb{E}[A(t)]$ exists we obtain that:*

$$\mathbb{E}[A(t)] = \int_0^t \mathbb{E}[dA(u)].$$

Proof. Let $S_i \sim \text{Gamma}(\psi, i)$, $X_i \sim f_i^\mu(x)$ and $Y_i(u) = \mathbb{1}\{S_i \leq u \leq T_i\}$ with $T_i = S_i + X_i$. We want to calculate:

$$\begin{aligned}\mathbb{E}\left[A(t)\right] &= \mathbb{E}\left[\int_0^t dA(u)\right] = \mathbb{E}\left[\int_0^t \sum_{i\geq 1} Y_i(u) h_i(u - S_i) du\right] \\ &= \int_0^\infty \int_0^t \sum_{i\geq 1} Y_i(u) h_i(u - S_i) f_i^\mu(x) du dx\end{aligned}$$

with $Y_i(u) = \mathbb{1}\{S_i \leq u\}\mathbb{1}\{X_i \geq u - S_i\}$, thus:

$$\mathbb{E}\left[A(t)\right] = \int_0^\infty \int_0^t \sum_{i\geq 1} \mathbb{1}\{S_i \leq u\}\mathbb{1}\{x \geq u - S_i\} h_i(u - S_i) f_i^\mu(x) du dx.$$

Now note that:

$$\sum_{i\geq 1} \mathbb{1}\{S_i \leq u\}\mathbb{1}\{X_i \geq u - S_i\} h_i(u - S_i) f_i^\mu(x) \tag{91}$$

is non-negative and assume that it is Borel measurable on $\mathbb{R} \times \mathbb{R}$. As indicator functions are always Borel measurable we only have to assume that the functions $h_i(x)$ and f_i^μ are Borel measurable. We assume that

$$\int_0^\infty \int_0^t \left| \sum_{i\geq 1} \mathbb{1}\{S_i \leq u\}\mathbb{1}\{X_i \geq u - S_i\} h_i(u - S_i) f_i^\mu(x) \right| du dx \tag{92}$$

exists, thus that the absolute value of expression (91) is integrable. Note that as all components of (91) are positive, this is the same as assuming that $\mathbb{E}\left[A(t)\right]$ exists. Then, using Fubini's theorem (or Tonelli's theorem as our integrands are non-negative) in section 23 of Bauer [42] we find that we can switch the order of integration:

$$\begin{aligned}\mathbb{E}\left[A(t)\right] &= \int_0^\infty \int_0^t \sum_{i\geq 1} \mathbb{1}\{S_i \leq u\}\mathbb{1}\{X_i \geq u - S_i\} h_i(u - S_i) f_i^\mu(x) du dx \\ &= \int_0^t \int_0^\infty \sum_{i\geq 1} \mathbb{1}\{S_i \leq u\}\mathbb{1}\{X_i \geq u - S_i\} h_i(u - S_i) f_i^\mu(x) dx du \\ &= \int_0^t \mathbb{E}\left[\sum_{i\geq 1} \mathbb{1}\{S_i \leq u\}\mathbb{1}\{X_i \geq u - S_i\} h_i(u - S_i)\right] du \\ &= \int_0^t \mathbb{E}\left[dA(u)\right].\end{aligned}$$

□

15.4 Standard simulation procedure

The standard simulation procedure employed in this thesis is described here. This procedure aims to restrict some theoretical quantity (such as the ARL or specificity) of the chart under the null hypothesis by choosing an appropriate control limit h. Afterwards the choice of control limit is evaluated for an out of control data set.

- Step 1: Generating a training (in control) data set with N hospitals.
  1. Choose null cumulative baseline hazard rate parametrically or determine from existing data (for example, using R package survival [17]).
  2. Generate patient arrival times in the required time frame using Poisson arrivals with rate ψ.
  3. (Optional) Bootstrap patient characteristics from data set.
  4. Determine (risk-adjusted) survival times for every patient using above chosen cumulative baseline hazard rate and section 15.1 with $\mu = 0$.
  5. Repeat 2-4 N times. Combine into single data set.
- Step 2: Determining a suitable control limit h.
  1. Determine a (parametric) cumulative baseline hazard rate using the generated in control data set (for example, using R package survival [17]). Optionally, use the cumulative hazard rate from step 1.
  2. Construct the charts on the training data set.
  3. Determine control limit h such that required restrictions (on sensitivity or ARL under the null) are met for the collection of the constructed charts.
- Step 3: Generate test (out of control) data set.
  1. Follow step 1 with $\mu > 0$ as required.
- Step 4: Evaluate charts on test data set
  1. Construct the charts on the test data with the control limit determined in step 2.
  2. Approximate required quantities (such as ARL, sensivity, specificity) from these charts.

15.5 Cox proportional hazards model graphical plots

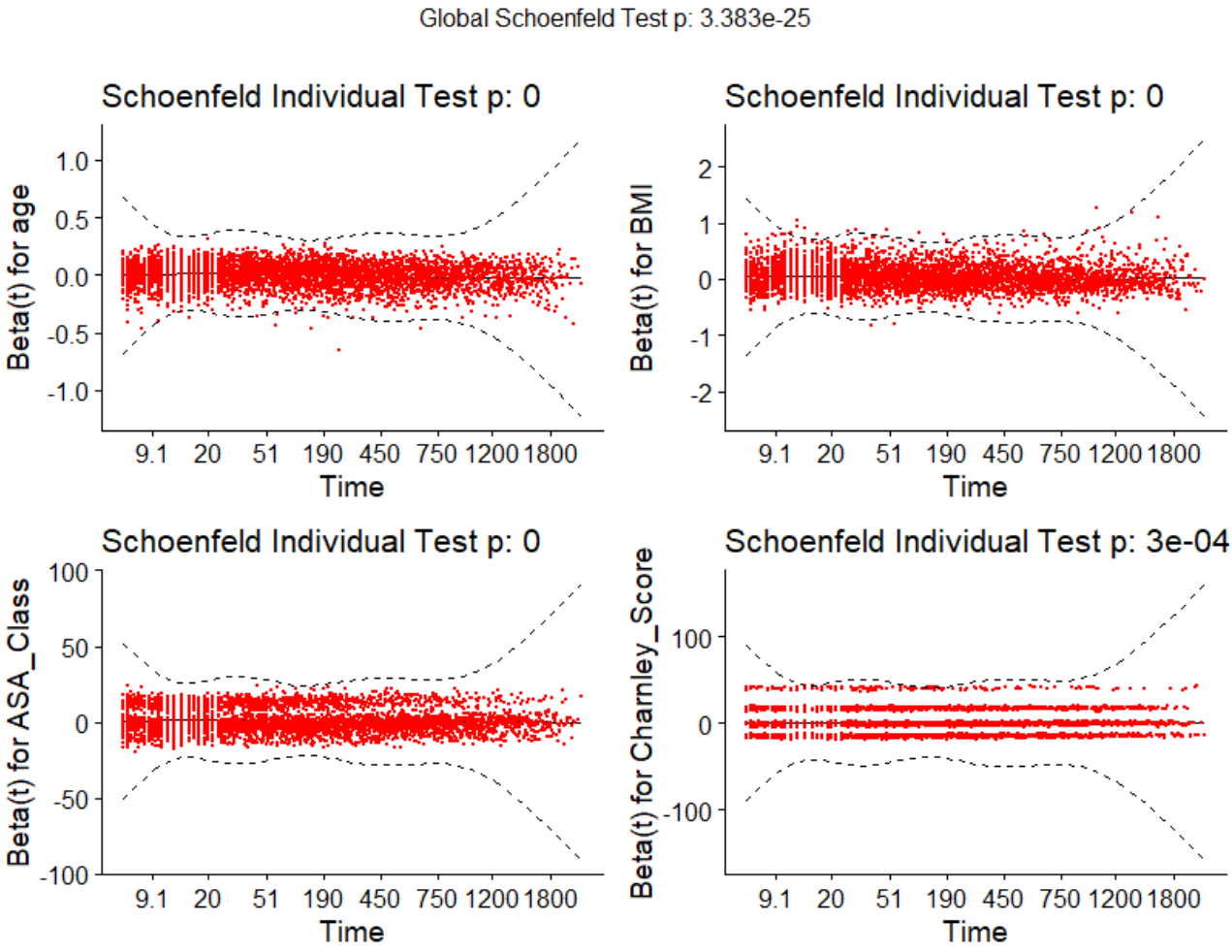


Figure 19: Scaled Schoenfeld residuals against time for the entirety of the LROI data set. Under the null hypothesis (PH assumption holds), the Schoenfeld residuals are independent of time. Any non-random pattern against time provides evidence for the violation of the null hypothesis.