Pseudo observations in multi-state models and CUSUM charts for monitoring outcomes of multi-center studies

by

Pinaki Biswas

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biostatistics)
in The University of Michigan
2007

Doctoral Committee:

Professor John D. Kalbfleisch, Chair
Professor Robert M. Merion
Professor Alexander Tsodikov
Assistant Professor Douglas E. Schaubel

Learning without thought is labor lost; thought without learning is perilous.

Confucius, *The Confucian Analect*,
Chinese philosopher and reformer (551 BC - 479 BC).

To my parents and my sister, who endured my absence in times of need with a cheerful smile.

ACKNOWLEDGEMENTS

It gives me great pleasure to be able to thank the people who have helped me write this dissertation. First and foremost, I would like to thank my advisor Professor John D. Kalbfleisch who has been extremely helpful, supportive and patient in guiding me at every stage in the dissertation. I would also like to extend my gratitude to Professor Robert M. Merion, Assistant Professor Douglas E. Schaubel and Professor Alexander Tsodikov for being on my committee and providing helpful inputs and comments. I am particularly thankful to Assistant Professor Schaubel for his careful reading of portions of the dissertation and offering constructive suggestions.

I would also like to thank Professor Susan Murray and Assistant Professor Adin-Cristian Andrei for helpful suggestions on pseudo observations and their applications. Thanks are also due to Mary K. Guidinger and Randall L. Webb for providing me much needed help and advice regarding application of the CUSUM procedure to kidney transplant data from the Scientific Registry of Transplant Recipients and organizing the results.

I would also like to thank Professor Moulinath Banerjee for informative and stimulating discussions on various theoretical and practical issues in statistics that helped broaden my knowledge. Thanks are also due to Professor Debashis Ghosh for helpful discussions and suggestions. I would also like to thank all my friends and colleagues for their encouragement and support during my graduate life in Ann Arbor.

I am grateful to the Department of Biostatistics for providing me financial support

throughout the length of my study, the faculty members of the Department for providing excellent courses, the staff members of the Department for helping me on several administrative issues and the University of Michigan for providing me excellent opportunities for learning.

I am also immensely grateful to the Indian Statistical Institute (ISI) for providing me, with full scholarship, an outstanding training and exposure to several areas of statistics, probability and mathematics. I am thankful to all my teachers at the ISI for providing excellent courses and encouraging me to pursue a career in research.

Finally and most importantly, I would like to thank my parents who have instilled in me the values of learning right from my childhood and have always encouraged me in the pursuit of knowledge. Without their blessings and sacrifice, this dissertation would not be possible.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

LIST OF APPENDICES

Appendix

CHAPTER I

Introduction

This dissertation essentially discusses two different problems. The first concerns an investigation of the use of pseudo observations in a multi-state event history model. In recent years, pseudo observations have been applied in different contexts as in Andersen et al. (2003), Andersen et al. (2004) and Klein and Andersen (2005) among several others. The second concerns the development and implementation of a CUmulative SUM (CUSUM) procedure in the context of monitoring outcomes of multi-center projects. The CUSUM procedure, introduced in Page (1954) and later discussed by Van Dobben de Bruyn (1968) and many others, has been used as a graphical sequential inspection scheme since the past several decades.

1.1 Multi-state models

A multi-state model can be described as one in which individuals can occupy any one of several states at an instant in time and can move between states as time progresses. It is common to describe the distribution of the times of transition from one state to another using a Cox proportional hazards model (Cox (1972)). Under a proportional hazards model, it is possible to write down an expression for the probability of a transition. Typically, this probability will depend on covariate values. It is important to be able to describe this relation through a simple and easily

interpretable model and a logistic regression model seems reasonable. In order to estimate the parameters of this model and obtain confidence intervals, Andersen et al. (2003) proposed using pseudo observations as the response in a binomial GLM. In this context, one can also construct confidence intervals using the bootstrap. Chapter II considers an illness-death model without recovery and discusses estimation of the probability of a transition from a healthy state to a diseased one under independent right-censoring. As in Andersen et al. (2003), a logistic model for this probability is considered and certain issues related to the implementation and performance of the pseudo observation approach are discussed and compared to bootstrap techniques. It becomes apparent through simulations and through theory from a simple example, that pseudo observations do not give correct answers when censoring is covariate-dependent.

1.2 CUSUM tests

CUSUM tests were originally developed as an extension of control chart techniques to monitor the quality of manufactured products. As the name suggests, they are based on cumulative sums of deviations from a fixed target and have proved to be generally sharper than control charts or Shewhart tests (Shewhart (1931)). As an illustration, let a process under inspection produce observations $x_1, x_2, x_3, \ldots$ in that order and let X be a target value associated with this process. The process is said to be in control if the mean of the $\{x_i\}$ stays close to this target value and potentially out of control if a systematic bias from the target is accumulated over time. Such a monitoring strategy is similar to a statistical test of hypothesis and therefore the idea of a sequential probability ratio test or SPRT was introduced in Wald (1947). Usually, the null hypothesis H_0 is that the process is in control with the alternative

H_1 being that it is biased away from the target. Then one can compute, at each stage of sampling, the ratio R of the likelihoods of obtaining the observed values under H_1 and H_0, respectively. If this ratio R is large, then H_0 is rejected; if it is small, then H_0 is accepted; and if it takes any intermediate value, the decision is delayed until further observations produce a higher or lower R. The difference between a CUSUM test and the SPRT is that the CUSUM test does not terminate when the H_0 is accepted, but rather restarts and continues until it leads to a rejection of H_0. The CUSUM test is then said to have signaled and the number of observations taken to achieve this is termed the Average Run Length or ARL. A CUSUM test may thus be seen as a sequence of SPRTs leading to rejection of H_0. Observe that the SPRT will always end by accepting or rejecting H_0, so it may be useful to design a CUSUM test in terms of achieving a particular ARL rather than probabilities of false positives or false negatives. Typically, it is desirable that the CUSUM should have a very large ARL when H_0 is true and a short ARL when H_1 is true. For the examples considered in this dissertation, only CUSUM tests having a one-sided alternative are of interest and accordingly a one-sided CUSUM (see Page (1954)) is developed and implemented.

1.3 Monitoring outcomes of multi-center studies

In multi-center projects involving an intervention, it is important to be able to monitor and provide outcome information in a timely manner to participating institutions or facilities. Such a monitoring activity can be useful in providing warning signals to the institutions and also in coordination of the project. Standard statistical techniques like risk-adjusted mortality and multivariate modeling can be used to identify performance changes at a national level but may be insensitive to smaller

persistent changes at the facility level. In the context of health-care applications, a one-sided risk-adjusted CUSUM has been discussed as a quality improvement scheme in Woodall et al. (2006) and particularly in the case of organ transplantation, by Axelrod et al. (2006), Poloniecki et al. (1998) and Steiner et al. (2001) among several others. In Axelrod et al. (2006) and Steiner et al. (2001), a risk-adjusted CUSUM based on a logistic model has been implemented in discrete time. The CUSUM needs to be adjusted for individual risk factors to account for institutional differences as can arise simply due to differences in the mix of patients. The implementation of such a procedure is at the institution of interest with the null hypothesis being that the rate of transplant failures at this institution is the same as that at the national or population level. It may be considered a cause for an alarm if there is evidence that the failure rate at this institution is higher than that at the national level, so the alternative is usually one-sided. In this dissertation, we develop and implement a risk-adjusted CUSUM using the Cox model (Cox (1972)) and based on an SPRT in continuous time. We also obtain approximate theoretical expressions for the ARL under certain assumptions, evaluate the performance through simulations and illustrate the procedure on a real data example. Both the simulations and the data example show that updating the CUSUM in continuous time leads to shorter detection times than the discrete versions.

It is helpful to keep in mind that one of the major objectives for implementing a CUSUM charting scheme is quality improvement. So a signal from the CUSUM should not be interpreted as a clinically important decline or improvement in clinical quality. Rather, the signal suggests that closer examination by the quality improvement team may be required. Mathematically, the CUSUM can be presented as a test of hypothesis problem, however, there is an important philosophical distinction

from the usual accept or reject outcome of a hypothesis test.

Chapter III develops and implements a one-sided CUSUM in continuous time and obtains approximate theoretical results on the ARL under the assumption that transplants occur according to a homogeneous Poisson process. The procedure is also demonstrated via a simulation study as well as on kidney transplant data from the Scientific Registry of Transplant Recipients (SRTR). In Chapter III, it is assumed that the report of an outcome is immediate. However, that is rarely the case in actual practice. Chapter IV proposes a CUSUM procedure when there is a random delay or lag involved between the time the outcome occurred and its reporting. In particular, we consider the case when the delays and the failure times are jointly independent conditional on the transplant times. The proposal is evaluated through simulation studies. The dissertation concludes with a short summary and discussion of future directions in Chapter V.

CHAPTER II

Pseudo obervations in multi-state models

2.1 Introduction

Multi-state models have long been used for modeling event history data and are discussed in various texts in the area including Andersen et al. (1993), Kalbfleisch and Prentice (2002) and Therneau and Grambsch (2000). It is traditional to specify multi-state models in terms of the transition intensity rates. Since the state occupancy probabilities are complicated functions of the transition intensities, even simple relationships between the covariates and transition intensities will generally result in complicated relationships between the covariates and state occupancy probabilities. In an attempt to address this issue, direct modeling of the state occupancy probabilities using General Linear Model methods has been suggested in Andersen et al. (2003). More specifically, the authors propose direct modeling of the dependence of the covariate effects on the state occupancy probabilities. To analyze this model, the authors devise a computationally simple approach utilizing pseudo observations generated from the Aalen-Johansen estimator (see, Aalen and Johansen (1978)).

Details of the proposal are as follows. If $\hat{\theta}$ is an unbiased estimator of θ, based on n observations, then the pseudo observations, as defined in Efron and Tibshirani

(1993), are,

$$\hat{\theta}_i = n\hat{\theta} - (n-1)\hat{\theta}_{(-i)}, \qquad i = 1, 2, \ldots, n \tag{2.1}$$

where $\hat{\theta}_{(-i)}$ is the value of $\hat{\theta}$ computed from all but the i-th observation. In Andersen et al. (2003), $\hat{\theta}_i$ is expressed in terms of the covariates $\mathbf{Z}_i$ as,

$$g(\hat{\theta}_i) \;=\; \boldsymbol{\beta}'\mathbf{Z}_i \tag{2.2}$$

where $\boldsymbol{\beta}$ is a vector of fixed unknown parameters and g is a known link function. The model is fitted using GEE techniques (Liang and Zeger (1986)). For the problem considered here, θ is an occupancy probability of interest and $\hat{\theta}$ is the Aalen-Johansen estimator of θ.

This article compares the above approach to a bootstrap technique in which percentile methods are used to obtain confidence intervals for the regression parameters in (2.2). The example we consider is a progressive illness-death model with independent right censoring. Also, a theoretical analysis of the effect of covariate-dependent censoring on estimates based on pseudo observations is presented in the context of a simple survival model with right censoring. Section 2.2 discusses the notation, assumptions, estimation procedures and the implementation of the bootstrap. Section 2.3 discusses a simulation of the illness-death model without recovery. Section 2.4 presents a data example from the SRTR to illustrate the approaches discussed in this article. Section 2.5 discusses a survival model with point censoring and a binary covariate and presents theoretical results for parameter estimates from the GLM based on pseudo observations. The article concludes with a discussion in Section 2.6.

2.2 A proposed GLM and its estimation

In this section, we introduce a multi-state model and describe an approximate GLM for a transition or occupancy probability, as suggested in Andersen et al. (2003), and discuss strategies for point and interval estimation of the parameters.

2.2.1 An approximate GLM

Let $\{X(t); t \geq 0\}$ denote the process for an individual where $X(t)$ takes values on a finite state space $\mathcal{S} = \{0, 1, 2, \ldots, m\}$. The transition probabilities are

$$P_{hj}(s,t) = \Pr(X(t) = j \mid X(s) = h, \mathcal{F}_{s-})$$

for time points $s < t$ and states $h,\ j \in \mathcal{S}$. Here, $\mathcal{F}_{s-} = \{X(t); 0 \leq u < s\}$. We assume that the process is Markov so that conditioning on the whole history $\mathcal{F}_{s-}$ gives no extra information beyond $X(s)$. For all $h \neq j,\ h, j \in \mathcal{S}$, the instantaneous transition intensities $\lambda_{hj}(t)$ are,

$$\lambda_{hj}(t) = \lim_{u \to 0^+} \frac{P_{hj}(t, t+u)}{u}.$$

We suppose that the process begins in state 1 and consider the occupancy probabilities for state h given by,

$$P_{1h}(t) = \Pr(X(t) = h). \tag{2.3}$$

To incorporate covariates into a multi-state model it is most natural to model the transition intensities, often as a Cox model (Cox (1972)), where

$$\lambda_{hj}(t; \mathbf{Z}) = \lambda_{0hj}(t) \exp(\boldsymbol{\gamma}_{hj}' \mathbf{Z}) \tag{2.4}$$

Though generally quite flexible and simple to interpret in terms of the transition intensities, (2.4) generally leads to a complicated description of the dependence of

the occupancy probability on $\mathbf{Z}$. In an attempt to address this problem, Andersen et al. (2003) considered an approximate model of the form

$$\text{logit}\{P_{1h}(t_l;\mathbf{Z})\} = \alpha_l + \boldsymbol{\beta}'\mathbf{Z}, \qquad l = 1, 2, \ldots, k. \tag{2.5}$$

where $\{t_1, t_2, \ldots, t_k\}$ are pre-determined time points and $P_{1h}(t_l;\mathbf{Z}) = \Pr(X(t) = h|\mathbf{Z})$.

In Andersen et al. (2003), it is noted that the model (2.5) would not hold exactly. Thus the authors suggest using a best fitting model of the form (2.5), though it is not explicitly described how that is to be done. Evidently, however, the idea, is to seek best values of $\{\alpha_1, \alpha_2, \ldots, \alpha_k, \boldsymbol{\beta}\}$ in the model (2.5) in order to be as close as possible to the underlying true model. Least squares would offer one approach; in this case the parameters are chosen to minimize

$$E_{\mathbf{Z}}\left[\sum_l \left\{\text{logit}(P_{1h}(t_l;\mathbf{Z})) - \alpha_l - \boldsymbol{\beta}'\mathbf{Z}\right\}^2\right].$$

If Z is a Bernoulli variate with $P(Z = 1) = \pi$, this yields,

$$\beta^{best} = \frac{1}{k}\sum_{l=1}^{k} \text{logit}\{P_{1h}(t_l; Z = 1)\} - \frac{1}{k}\sum_{l=1}^{k} \text{logit}\{P_{1h}(t_l; Z = 0)\} \tag{2.6}$$

$$\alpha_l^{best} = \pi\text{logit}\{P_{1h}(t_l; Z = 1)\} + (1 - \pi)\text{logit}\{P_{1h}(t_l; Z = 0)\} - \pi\beta^{best} \tag{2.7}$$

Further, if $k = 1$, there is an exact relation for the occupancy probability. Specifically,

$$\text{logit}\{P_{1h}(t_1; Z)\} = \alpha_1^{best} + \beta^{best} Z, \qquad Z = 0, 1 \tag{2.8}$$

Note that this is just a special case of (2.5), so β^{best} and α_1^{best} in this case are obtained by substituting $k = 1$ in (2.6) and (2.7), respectively. More generally, with $k > 1$ and/or non-binary Z, (2.6) and (2.7) offer only an approximate solution. This exact solution when $k = 1$ is useful in comparing the various approaches, since all proposed models are then correct and so on the same footing.

2.2.2 Estimation

In order to estimate the model parameters $\{\alpha_1^{best}, \alpha_2^{best}, \ldots, \alpha_k^{best}, \beta^{best}\}$ as in (2.6) and (2.7), we need estimates for the probabilities $P_{1h}(t; Z)$. For simplicity and clairity, we restrict attention to binary $Z = 0, 1$ for $k > 1$ and $k = 1$. We now examine each of the three approaches to estimation in the following three subsections.

Pseudo observations

For given t, we estimate $\theta = P_{1h}(t) = E[P_{1h}(t; Z)]$ nonparametrically with the Aalen-Johansen estimator $\hat{\theta}$ based on the data on all n sampled individuals. We then omit the i-th individual to compute $\hat{\theta}_{(-i)}$ for $i = 1, 2, \ldots, n$. The pseudo observation $\hat{\theta}_i$ as defined in (2.1) is viewed as an approximate observation on $P_{1h}(t; Z_i)$. Note that we do not use any covariate information for forming the pseudo observations.

An immediate difficulty is that the pseudo observation need not lie in $[0, 1]$. One approach is to set negative pseudo observations to 0 and those larger than 1 to 1. These truncated pseudo observations are then used as the response in a binomial GLM with a logit link to obtain estimates for β^{best} in (2.6). Note that for $k > 1$ in (2.5), $\boldsymbol{\theta}$ is the vector of k occupancy probabilities and so we have a k-vector of pseudo observations for each individual. Although we used an independence structure for the working covariance matrix in the GEE, other structures may also be used (Klein and Andersen (2005)). Following Andersen et al. (2003), confidence intervals are formed using the assumed asymptotic normality of $\hat{\beta}$ with standard errors obtained from the sandwich formula.

An alternative to truncation is to solve the logistic regression equations (Andersen

and Klein (2007), Klein and Andersen (2005)),

$$\sum_{i=1}^{n}(\hat{\theta}_{ij} - p_{ij})p_{ij}(1-p_{ij}) \;=\; 0, \qquad j = 1, 2, \ldots, k \tag{2.9}$$

$$\sum_{i=1}^{n} Z_i \sum_{j=1}^{k}(\hat{\theta}_{ij} - p_{ij})p_{ij}(1-p_{ij}) \;=\; 0 \tag{2.10}$$

where $\text{logit}(p_{ij}) = \alpha_j^{best} + \beta^{best} Z_i$. For $k = 1$, it is immediate that,

$$\hat{\beta}^{best} \;=\; \text{logit}\left(\frac{\sum_i \hat{\theta}_i Z_i}{\sum_i Z_i}\right) - \text{logit}\left(\frac{\sum_i \hat{\theta}_i (1-Z_i)}{\sum_i (1-Z_i)}\right) \tag{2.11}$$

Standard errors are again computed using the sandwich estimator. In this, we ignore the fact that the pseudo observations are estimates of probabilities and so arguably should lie in $[0, 1]$; the approximate linearization giving rise to the pseudo observations suggests that this may be the more appropriate approach.

In the case of a general Markov process discussed earlier, the Aalen-Johansen (AJ) estimate for the transition probability matrix $\mathbf{P}(s, t)$ can be written as,

$$\hat{\mathbf{P}}(s,t) \;=\; \prod_{i: s \le T_i \le t} (\mathbf{I} + d\hat{\mathbf{A}}(T_i))$$

where $\{s \le T_i \le t\}$ are the transition times (between any two states) that are observed to lie between s and t. The matrix $\hat{\mathbf{A}}(t) = (\hat{A}_{hj}(t))$ is,

$$\hat{A}_{hj}(t) \;=\; \int_0^t I(Y_h(u) > 0)\frac{dN_{hj}(u)}{Y_h(u)}.$$

Here, $N_{hj}(t)$ counts the number of observed direct transitions from state h to j in $[0, t]$, and has intensity process $\lambda_{hj}(t)Y_h(t)$, where $Y_h(t)$ is the number of individuals observed to be in state h just prior to time t. Thus, the computation of the AJ estimator involves a finite number of matrix multiplications.

A nonparametric estimator

As a second estimator, we consider the AJ estimator stratified on the levels of Z. Obviously, this approach is applicable only with discrete covariates taking on

relatively few distinct values. This method is used here mainly as a nonparametric yardstick and is not proposed as a possible alternative. The estimates of α_j^{best} and β^{best} are obtained by solving the equations (2.9) and (2.10) with $\hat{\theta}_{ij}$ replaced with the corresponding AJ estimates.

An estimator based on the Cox model

Under the assumption (2.4) for the transition intensity functions, estimation of γ_{hj} can be based on the partial likelihood and $\Lambda_{0hj}(t)$ can be estimated with the Nelson-Aalen estimator; (see, for example, Kalbfleisch and Prentice (2002), Andersen et al. (1993)). As demonstrated for a simple example in the next section, these estimates can be substituted into an integral representation of $P_{1h}(t_j; Z_i)$ to obtain an estimate. More generally, an estimate of $P_{1h}(s, t; Z)$ is available through substitution of estimated quantities into its product integral representation. Finally, estimates of α_j^{best} and β^{best} are obtained by substituting Cox model estimates of $P_{1h}(t_j; Z_i)$ for $\hat{\theta}_{ij}$ in (2.9) and (2.10).

2.2.3 Confidence intervals using the bootstrap

To obtain confidence intervals for the AJ and Cox model approaches we use a bootstrap approach and obtain $\hat{\beta}^{best*}$ for each resampled data set. Resampling introduces ties in the data, even if none are present initially, and we used Efron's approximation (Efron (1977)). The lower and upper limits for an approximate $100(1-\gamma)\%$ confidence interval are obtained as the $\gamma/2$ and $(1-\gamma/2)$ quantiles of the bootstrap distribution of the estimate of β^{best}.

2.3 Simulation study

In this section, we evaluate the three methods described above in a simple multistate model with independent right censoring.

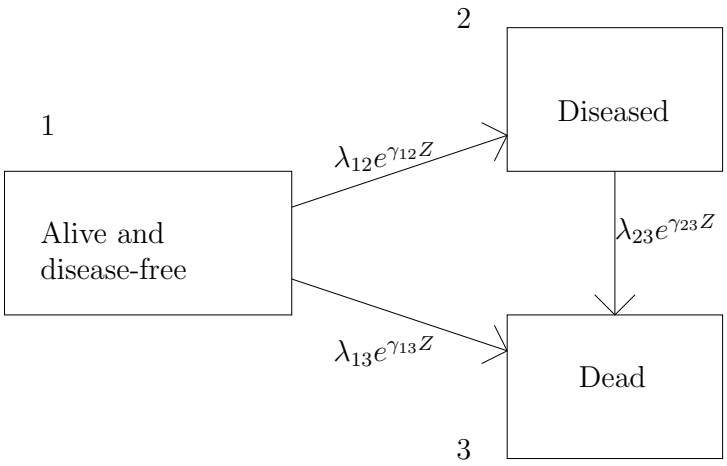

Figure 2.1: An illness-death model without recovery.

2.3.1 An illness-death model

We consider the progressive illness-death model with 3 states as illustrated in Figure 2.1. For the simulations, we used a binary covariate ($Z = 0, 1$) and constant baseline intensities for each of the transitions as indicated in the figure. We considered estimation of the probability that an individual who is disease-free at time zero is alive and in the diseased state at time t. This is,

$$P_{12}(t;Z) = \int_0^t P_{11}(0,u-;Z)\lambda_{12}(u;Z)P_{22}(u+,t;Z)du$$

where,

$$\begin{aligned} P_{11}(0,t;Z) &= \exp\left\{-\int_0^t (\lambda_{12}(u;Z)+\lambda_{13}(u;Z))du\right\},\ t>0 \ \text{ and} \\ P_{22}(s,t;Z) &= \exp\left\{-\int_s^t \lambda_{23}(u;Z)du\right\},\ s<t. \end{aligned}$$

Under assumptions of constant baseline intensities, we find,

$$\begin{aligned} P_{12}(t;Z) = &\frac{\lambda_{12}e^{\gamma_{12}Z}\exp\left\{-\lambda_{23}e^{\gamma_{23}Z}t\right\}}{\lambda_{12}e^{\gamma_{12}Z}+\lambda_{13}e^{\gamma_{13}Z}-\lambda_{23}e^{\gamma_{23}Z}} \times [1- \\ &\exp\left\{-(\lambda_{12}e^{\gamma_{12}Z}+\lambda_{13}e^{\gamma_{13}Z}-\lambda_{23}e^{\gamma_{23}Z})t\right\}], \end{aligned}$$

which is a complicated expression even for this simple model. It is therefore not an easy task in general to describe the nature of the dependence of $P_{12}(t;Z)$ on Z in a simple manner.

To illustrate the methods described in Section 2.2, we carried out a simulation study with constant baseline censoring rates that were allowed to depend on Z as well as the state currently occupied. Thus, $\lambda_{jC}e^{\gamma_{jC}Z}$ denotes the hazard rate of censoring for an individual in state j, $j = 1, 2$. The overall proportion of censoring was controlled by fixing the unconditional probability of being censored at a desired fraction α.

The parameter values for the simulation were chosen as, $\pi = 0.5, \lambda_{12} = 1, \lambda_{13} = 0.5, \lambda_{23} = 2$ and $\gamma_{12} = \gamma_{13} = \gamma_{23} = \log 2$. For censoring, we considered $(\gamma_{1C}, \gamma_{2C}) = (0, 0)$ or $(\gamma_{1C}, \gamma_{2C}) = (\log 5, \log 2)$. We considered the cases $\lambda_{1C} = \lambda_{2C} = \lambda$ and $\lambda_{1C} = \lambda$, with $\lambda_{2C} = 0.5$. The first setting is similar to a controlled trial where drop-out rates are regulated and the second is similar to an observational study where healthy individuals are more likely to drop out (censored) than are sick individuals. In each setting, the value of λ was chosen to ensure a given fraction α (= 0.0, 0.25, 0.50 and 0.75) of overall censoring.

Figure 2.2 displays the true model probability $P_{12}(t;Z)$ for $Z = 0$ and 1. It is evident from the graph that the curves vary considerably with time and are far from parallel and so β^{best} from model (2.5) is measuring only an average effect over time. Table 2.1 shows the probability estimates under a certain simulation setting using the Aalen-Johansen (AJ) and Cox methods. Both the procedures appear to be quite accurate. The estimates in Table 2.1 are based on 1000 Monte Carlo replications.

We report on the estimation of regression parameters for the cases $k > 1$ and $k = 1$. In the first (CASE I), we chose $k = 4$ and time points $\{0.10, 0.15, 0.20, 0.25\}$;

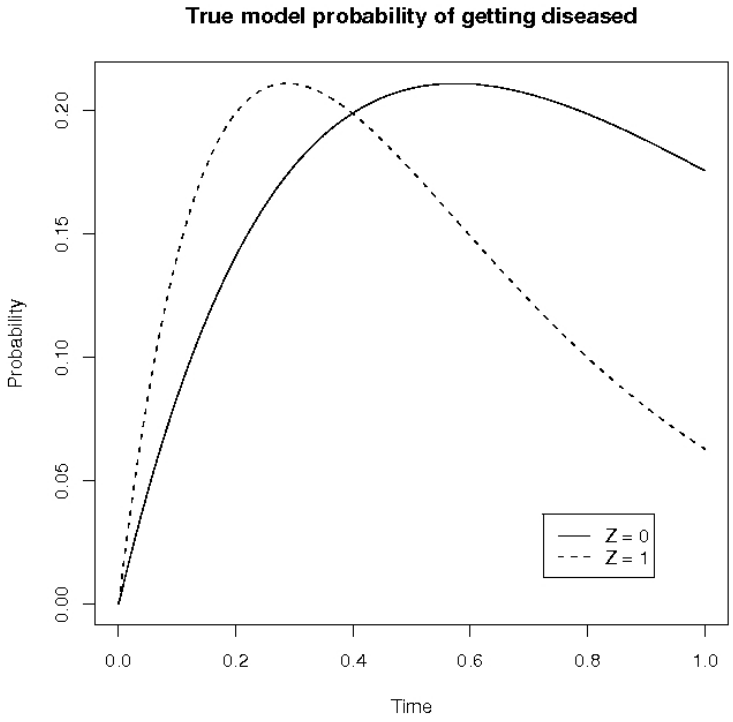

Figure 2.2: True probabilities $P_{12}(t;Z)$ from the illness-death model for $Z=0$ and $Z=1$.

Table 2.1: True and estimated probabilities with $\alpha=0.5, n=300, \lambda_{1C}=\lambda_{2C}, \gamma_{1C}=\gamma_{2C}=0$.

Time	True probability $P_{12}(t;Z=0)$	AJ estimator $\hat{P}_{12}^{\mathrm{AJ}}(t;Z=0)$	Cox estimator $\hat{P}_{12}^{\mathrm{Cox}}(t;Z=0)$
0.1	0.084	0.083	0.082
0.2	0.141	0.140	0.140
0.3	0.178	0.176	0.175
0.4	0.199	0.197	0.196
0.5	0.209	0.208	0.206
0.6	0.211	0.208	0.207
0.7	0.207	0.206	0.204
0.8	0.199	0.197	0.197
0.9	0.188	0.185	0.190
1.0	0.176	0.177	0.190

Table 2.2: Estimates of β^{best}, coverage probabilities (CP) of nominal 95% interval and average interval length (AIL) for the illness-death model when $n = 300$ under CASE I ($k = 4$). The true β^{best} is 0.454.

No censoring

	AJ	Cox	PO-Tr	PO
$\hat{\beta}$	0.451	0.449	0.437	0.440
CP	94.2	94.0	94.4	94.3
AIL	1.065	0.650	0.969	0.970

$\lambda_{1C} = \lambda_{2C} = \lambda$

		$\gamma_{1C} = \gamma_{2C} = 0$				$\gamma_{1C} = \log 5, \gamma_{2C} = \log 2$			
α		AJ	Cox	PO-Tr	PO	AJ	Cox	PO-Tr	PO
	$\hat{\beta}$	0.442	0.445	0.435	0.442	0.462	0.459	0.444	0.458
0.25	CP	94.5	95.0	94.8	94.8	93.6	94.9	95.0	95.0
	AIL	1.094	0.711	1.005	1.005	1.108	0.751	1.025	1.026
	$\hat{\beta}$	0.449	0.455	0.431	0.446	0.446	0.452	0.469	0.529
0.5	CP	93.8	94.3	94.3	94.6	94.6	93.8	95.4	94.6
	AIL	1.162	0.824	1.074	1.069	1.197	0.920	1.133	1.138
	$\hat{\beta}$	0.453	0.449	0.442	0.469	0.454	0.454	0.437	0.770
0.75	CP	94.6	94.6	94.4	95.0	93.2	94.2	97.8	97.6
	AIL	1.377	1.097	1.313	1.282	1.522	1.299	1.426	1.589

$\lambda_{1C} = \lambda, \lambda_{2C} = 0.5$

		$\gamma_{1C} = \gamma_{2C} = 0$				$\gamma_{1C} = \log 5, \gamma_{2C} = \log 2$			
α		AJ	C	PO-Tr	PO	AJ	C	PO-Tr	PO
	$\hat{\beta}$	0.457	0.453	0.429	0.436	0.450	0.452	0.433	0.444
0.25	CP	93.6	95.4	96.1	96.0	95.3	95.1	94.7	94.8
	AIL	1.095	0.708	1.000	1.000	1.092	0.717	1.006	1.007
	$\hat{\beta}$	0.464	0.464	0.423	0.442	0.451	0.452	0.452	0.519
0.5	CP	94.4	94.3	94.8	95.0	94.7	95.3	95.5	95.2
	AIL	1.179	0.866	1.098	1.090	1.202	0.933	1.136	1.141
	$\hat{\beta}$	0.453	0.455	0.435	0.463	0.423	0.427	0.374	0.922
0.75	CP	94.3	94.8	95.6	96.2	93.2	93.8	96.7	97.2
	AIL	1.508	1.281	1.444	1.391	1.732	1.463	1.540	1.995

over this region the curves are almost parallel (see Figure 2.2). In the second (CASE II), we took $k = 1$ with $t_1 = 0.4$; for this case, the model for the occupancy probability holds exactly.

Simulations were conducted using $B = 2000$ bootstrap samples and a sample of size $n = 300$. Point estimates of β^{best} were obtained based on the three methods of Section 2.2.2. Confidence intervals for the AJ and the Cox estimators were obtained from the bootstrap, whereas the sandwich estimator was used for the pseudo observation approach. The procedure is heavily computational.

Table 2.2 displays simulation results for CASE I with $k = 4$ time points for which $\beta^{best} = 0.454$. The table displays the estimate of β^{best}, followed by the estimated coverage probability and the estimated average length of the interval with α denoting the fraction of censoring. In the tables, PO-Tr represents the pseudo observation approach with truncation whereas PO represents the same without truncation as discussed in Section 2.2.2, and AJ and Cox denote the methods based on Aalen-Johansen estimators and the Cox model, respectively.

For the case of no censoring ($\alpha = 0$), all four methods have good coverage properties although the PO estimators show some small bias as well as larger intervals. The PO method without truncation encounters difficulty with high (75%) censoring, where for around 3% of the time, the estimating equations (2.9) and (2.10) in Section 2.2.2 admit no solution in the parameter space. This fraction goes up with smaller sample sizes (about 12% for $n = 100$) or with higher censoring. This problem becomes more serious in the tail region ($t > 0.5$) owing to the censoring. To get around this, we left such samples out of the calculation and also considered CASE I ($k = 4$) with early time points for which the untruncated PO method rarely fails, at least for the lower censoring rates. When the censoring is allowed to depend on Z, the PO methods show substantial bias and break down rather quickly. The approach based on the Cox model does not suffer this problem and works well in all cases considered.

Finally, Table 2.3 displays simulation results for CASE II ($k = 1$) with $t_1 = 0.4$. As noted earlier, this model is consistent with both the Cox model and the logistic model for $P_{12}(t; Z)$, thereby providing a common footing for comparison. The value of β^{best} is about -0.002 reflecting the fact that $t_1 = 0.4$ is very close to the crossing point of the probability curves in Figure 2.2. Although the coverage probabilities for

Table 2.3: Estimates of β^{best}, coverage probabilities (CP) of nominal 95% interval and average interval length (AIL) for the illness-death model when $n = 300$ under CASE II ($k = 1$). The true β^{best} is -0.002.

No censoring

	AJ	Cox	PO-Tr	PO
$\hat{\beta}$	0.005	-0.001	0.005	0.012
CP	94.4	95.5	91.4	95.1
AIL	1.123	0.716	0.756	1.015

$\lambda_{1C} = \lambda_{2C} = \lambda$

		$\gamma_{1C} = \gamma_{2C} = 0$				$\gamma_{1C} = \log 5, \gamma_{2C} = \log 2$			
α		AJ	Cox	PO-Tr	PO	AJ	Cox	PO-Tr	PO
	$\hat{\beta}$	-0.001	-0.011	0.138	-0.002	0.012	0.002	-0.202	0.042
0.25	CP	95.3	95.0	90.6	95.0	94.5	94.7	86.9	96.8
	AIL	1.273	0.797	0.748	1.153	1.329	0.833	0.696	1.138
	$\hat{\beta}$	-0.141	0.007	-0.210	-0.018	0.017	0.004	-0.189	0.052
0.5	CP	96.0	94.9	80.9	95.4	91.9	95.6	82.9	97.7
	AIL	1.454	0.952	0.756	1.449	1.849	1.050	0.650	1.378
	$\hat{\beta}$	-0.014	0.028	-0.341	-0.051	0.036	-0.007	-0.241	-0.006
0.75	CP	94.4	92.2	68.4	98.3	93.6	93.8	67.5	99.4
	AIL	2.743	1.381	0.860	13.610	2.782	1.575	0.592	2.049

$\lambda_{1C} = \lambda, \lambda_{2C} = 0.5$

		$\gamma_{1C} = \gamma_{2C} = 0$				$\gamma_{1C} = \log 5, \gamma_{2C} = \log 2$			
α		AJ	Cox	PO-Tr	PO	AJ	Cox	PO-Tr	PO
	$\hat{\beta}$	-0.011	-0.004	0.190	0.001	-0.009	-0.006	-0.174	0.076
0.25	CP	92.0	95.3	85.3	95.5	95.6	94.9	83.1	95.7
	AIL	1.281	0.794	0.752	1.145	1.360	0.814	0.712	1.136
	$\hat{\beta}$	-0.139	0.003	-0.209	0.025	-0.145	-0.012	-0.194	0.065
0.5	CP	93.0	96.2	79.1	95.9	92.2	95.3	74.9	97.3
	AIL	1.373	0.974	0.692	1.445	1.601	1.054	0.656	1.380
	$\hat{\beta}$	-0.260	0.017	-0.225	0.015	-0.208	-0.039	-0.252	-0.126
0.75	CP	92.1	94.7	71.1	97.7	93.1	94.8	68.2	99.1
	AIL	2.177	1.444	0.644	2.669	2.227	1.642	0.576	2.108

all the methods are reasonable, the Cox method yields the shortest intervals. Again, the PO methods break down as the degree of censoring increases, and they exhibit bias when the censoring is allowed to depend on Z. In all the scenarios, the PO method without truncation outperforms the method with truncation incorporated. For CASE II, it is interesting to look at the efficiency of the PO estimator relative to the Cox estimators by considering the ratio of average interval widths. From Table 2.3 we see that this ranges from 65% to 78%. Also, it may be useful to consider a robust measure of interval width like the median to discount extreme values (as in the case $\lambda_{1C} = \lambda_{2C} = \lambda, \gamma_{1C} = \gamma_{2C} = 0.0, \alpha = 0.75$).

Theoretical results behind the numbers in Tables 2.2 and 2.3 are limited. Our primary objective was to investigate how the pattern of censoring may affect the performance of the PO method. We do however, pursue some theoretical insights into the PO method for a simple example in Section 2.5.

2.4 An illustration

In the recent past, a discussion group was formed to assess the use of pseudo observations for estimating mean residual life of subjects in the Scientific Registry of Transplant Recipients (SRTR) and this article is driven by certain issues raised in the group. In relation to this, we consider an illustration of the above estimation methods on data from the SRTR. Our data include information on 8798 individuals who had been waitlisted for a kidney transplant in the year 1987. The data include the date the individual was waitlisted, the date of transplant (if transplant received), the date of removal from the waitlist (either due to transplant, death or other causes) and the date of death if the individual had died. We chose January 1st 1989 as the time to end of follow-up and so all events occurring after this date were considered

censored. Individuals removed from the waitlist for reasons other than transplant were also considered censored on the waitlist. Of the 8798 individuals waitlisted, 3872 received a kidney transplant, 543 died on the waitlist. Of the 3872 individuals who received a transplant, 252 individuals died. For each individual, age at listing was noted among various other measurements.

The progressive illness-death model (Section 2.3) was employed, with being on the waitlist corresponding to state 1, transplanted and alive corresponding to state 2 and state 3 corresponding to death. The probability of interest is the probability of being transplanted and alive at any time. We carried out the analysis using age as the covariate. Age was coded as binary, with 0 indicating younger than 18 years and 1 indicating at least 18 years. We fitted the GLM (2.5) with $k > 1$ and $k = 1$ to the data. For fitting the Cox model the following forms were assumed for the three transition intensities.

$$\begin{aligned} \lambda_{12}(t; Z) &= \lambda_{012}(t) \exp\{\gamma_{12} Z\} \\ \lambda_{13}(t; Z) &= \lambda_{013}(t) \exp\{\gamma_{13} Z\} \\ \lambda_{23}(t; Z) &= \lambda_{023}(t) \exp\{\gamma_{23} Z\} \end{aligned}$$

Table 2.4: Table showing parameter estimates and confidence intervals with age category as the covariate.

	$k = 15$			$k = 1$		
	AJ	Cox	PO	AJ	Cox	PO
$\hat{\beta}^{best}$	-0.35	-0.29	-0.31	-0.29	-0.28	-0.22
CI	(-0.52, -0.14)	(-0.45, -0.12)	(-0.49, -0.12)	(-0.49, -0.06)	(-0.45, -0.12)	(-0.43, -0.01)

Time was measured in years and we fit model (2.5) using $k = 15$ points as $\{i/10 : i = 1, 2, \ldots, 15\}$ and $k = 1$ with $t_1 = 0.8$. Table 2.4 shows the point and interval estimates for β^{best} from the three methods AJ, Cox and PO using age category as

the covariate. The Cox model yields the shortest intervals, particularly for $k = 1$, when all three methods are comparable. Also, for $k = 1$, the point estimates look similar with the PO estimator in slight disagreement. For $k = 15$, the AJ and PO estimators are close.

The analysis shows that the probability of getting a transplant and remaining alive increases with time on the waitlist. Figure 2.3 shows probability estimates from the AJ and the Cox approaches. The estimated curves demonstrate that younger individuals are more likely to get transplanted than adult individuals. The estimates from the AJ and Cox methods are close for $Z = 1$ (Age ≥ 18) and in slight disagreement for $Z = 0$ (Age < 18). It is useful to remember here that the risk-adjustment may not be adequate and that this discussion is merely for illustration.

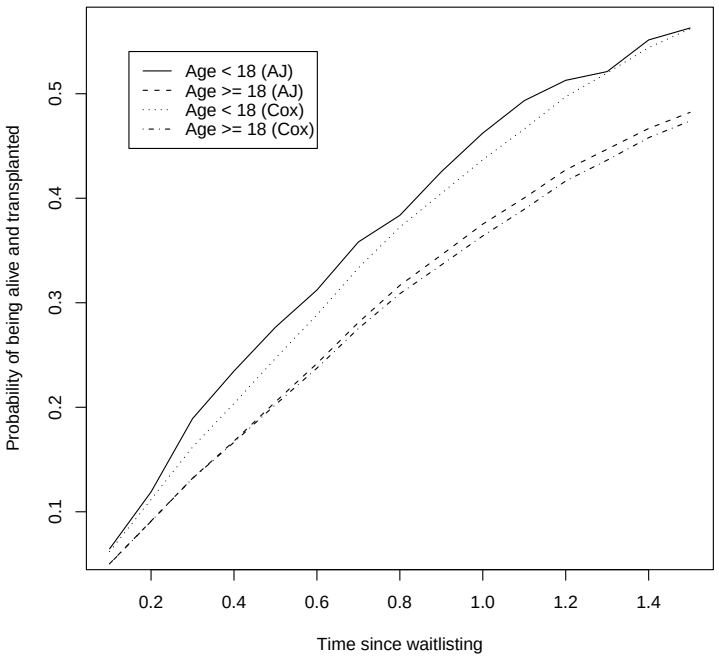

Figure 2.3: Nonparametric (AJ) and Cox estimate of the probability of being transplanted and alive, by age group.

2.5 A simple and instructive example

It is useful to investigate the performance of the PO method when the level of censoring is covariate-dependent. It is difficult, however, to establish general results. In this section, we consider a simple survival model with independent right censoring and a binary covariate as before. We derive expressions for the pseudo observations, obtain an exact expression for the parameter estimates in (2.11) and show that these estimates can be inconsistent depending on the pattern of censoring.

Let $\{(X_i, \Delta_i, Z_i)\}_{i=1}^n$ be the observed data with $X_i = \min(T_i, C_i)$ and $\Delta_i = I(T_i \leq C_i)$ with T_i and C_i being the failure and censoring times, respectively. Assume that the censoring times C_i are such that $P(C_i = C) = 1$ for every i, where $C \geq 0$ is a known constant.

Now let Z_i be binary with $P(Z_i = 1) = \pi$. Let n_j be the number of individuals with $Z_i = j,\ j = 0, 1$. Assume also, that a fraction α_j of individuals are randomly chosen for censoring in group $Z_i = j$. We consider a single time point t_1 ($k = 1$) such that $t_1 > C$.

For $j = 0, 1$, let u_{1j} be the number observed to fail before C, R_j be the number censored at C and u_{2j} be the number of observed failures between C and t_1. Also, let $n = n_0 + n_1,\ u_1 = u_{10} + u_{11},\ R = R_0 + R_1$ and $u_2 = u_{20} + u_{21}$ reflect the overall numbers. The pseudo observations $\hat{\theta}_i$, based on the Kaplan-Meier estimator of $\theta = S(t_1) = P(T_i > t_1)$, are

$$\hat{\theta}_i \;=\; \begin{cases} 0 & i \leq u_1, \\ 1 - \frac{u_2}{n-R} & u_1 < i \leq u_1 + R, \\ -\frac{R(n-R-u_2)}{(n-R)(n-R-1)} & R < i \leq R + u_2, \\ \frac{n(n-R-u_2)}{n-R} - \frac{(n-1)(n-R-u_2-1)}{n-R-1} & R + u_2 < i \leq n \end{cases}$$

where we have, without loss of generality, assumed that the observations are

in ascending order of X_i. On solving the estimating equations (2.9) and (2.10) in Section 2.2.2, the point estimator of $p_0 = P(T_i > t_1|Z_i = 0)$ based on $\hat{\theta}_i$ is $[\sum_{i=1}^{n} \hat{\theta}_i(1 - Z_i)]/[\sum_{i=1}^{n}(1 - Z_i)]$. After some simplification, we obtain,

$$\begin{aligned} n_0\hat{p}_0 &= R_0(1 - \frac{u_2}{n - u_1 - R}) - u_{20}\frac{n - u_1 - 1}{n - u_1 - R - 1} \\ &\quad +(n_0 - R_0 - u_{10})\left[1 + \frac{u_2 R}{(n - u_1 - R)(n - u_1 - R - 1)}\right] \end{aligned}$$

It is interesting to investigate the probability limit of $\hat{p}_0$ as $n \to \infty$. We know that $R_j \sim$ Bin(k_j, λ_j), where $\lambda_j = S_j(C) = P(T > C|Z = j)$ and $k_j = \alpha_j n_j$. Also, $n_1 \sim$ Bin (n, π). Furthermore, $u_{1j} \sim$ Bin$(n_j, 1-\lambda_j)$ and $u_{2j} \sim$ Bin$(n_j, (1-\alpha_j)(\lambda_j - p_j))$, where $p_j = \Pr(T > t_1|Z = j)$. Under the above distributional assumptions, we obtain, after some simplification,

$$\begin{aligned} \hat{p}_0 &\xrightarrow{p} \lambda_0 - \frac{(1-\alpha_0)(\lambda_0 - p_0)\left[(1-\pi)\lambda_0 + \pi\lambda_1\right]}{(1-\pi)\lambda_0(1-\alpha_0) + \pi\lambda_1(1-\alpha_1)} \\ &\quad +\pi\lambda_0\lambda_1(\alpha_1 - \alpha_0)\frac{\left[(1-\pi)(1-\alpha_0)(\lambda_0 - p_0) + \pi(1-\alpha_1)(\lambda_1 - p_1)\right]}{\left[(1-\pi)\lambda_0(1-\alpha_0) + \pi\lambda_1(1-\alpha_1)\right]^2} \end{aligned}$$

From the above expression, it is easy to see that when $\alpha_0 = \alpha_1$, then the limit is p_0. Other than this case, the limit is typically different from p_0 and therefore $\hat{p}_0$ is inconsistent. The calculations and results for $p_1 = P(T_i > t_1|Z_i = 1)$ are similar. In order to see this in practice, we conducted a simulation study with $n = 5000,\ t_1 = 0.5,\ C = 0.3$ and 10,000 replications. The failure time distribution was taken as exponential with rate $2^j,\ j = 0, 1$. We also chose $\pi = P(Z = 1) = 1 - P(Z = 0) = 0.5$.

Table 2.5 shows coverage probabilities for confidence intervals based on pseudo observations under different choices of α_0 and α_1. The objective is estimation of $\beta =$ logit $p_1 -$ logit p_0. Observe that when α_0 is large and α_1 is small, there is more bias in $\hat{p}_0$ than $\hat{p}_1$ and so this effect is carried over to $\hat{\beta}$. The effect is opposite when α_0

Table 2.5: Parameter estimates, average interval lengths and coverage probability (CP) for the simple model considered at $t_1 = 0.5$ when $n = 5000$ and $C = 0.3$. Target coverage is 95%.

			Pseudo observations				Kaplan-Meier				Efficiency
α_0	α_1	α	$\hat{p}_0$	$\hat{p}_1$	$\hat{\beta}$	CP	$\hat{p}_0^{KM}$	$\hat{p}_1^{KM}$	$\hat{\beta}^{KM}$	CP^{KM}	$\frac{MSE^{KM}}{MSE^{PO}}$
0.10	0.60	0.20	0.615	0.393	-0.902	81	0.606	0.368	-0.975	95	0.55
0.10	0.70	0.23	0.615	0.403	-0.862	60	0.606	0.368	-0.972	95	0.35
0.10	0.80	0.26	0.614	0.415	-0.811	27	0.607	0.367	-0.978	95	0.24
0.00	0.00	0.00	0.607	0.368	-0.975	96	0.607	0.368	-0.975	96	1.00
0.40	0.40	0.26	0.607	0.368	-0.975	95	0.607	0.368	-0.975	95	1.00
0.60	0.60	0.39	0.606	0.368	-0.976	95	0.606	0.368	-0.976	95	1.00
0.80	0.80	0.52	0.607	0.368	-0.974	95	0.607	0.368	-0.974	95	1.00
0.60	0.10	0.25	0.582	0.353	-0.937	93	0.607	0.368	-0.975	95	0.83
0.70	0.10	0.29	0.572	0.352	-0.898	83	0.606	0.368	-0.975	95	0.53
0.80	0.10	0.32	0.556	0.353	-0.831	52	0.606	0.368	-0.974	95	0.27

True values $p_0 = 0.606, p_1 = 0.368, \beta^{best} = -0.974$.

and α_1 are switched. When α_0 and α_1 are close ($\alpha_0 \approx \alpha_1$), then the bias is small (from the limit above) in both directions and so the effects get somewhat compensated in $\hat{\beta}$. Also figuring in the same table are estimates and confidence intervals based on the usual Kaplan-Meier estimator. Estimates for p_0, p_1 are obtained after stratifying for $Z = 0, 1$ and confidence intervals are based on standard errors obtained from Greenwood's formula. This procedure performs quite well, for any censoring plan, as expected. Table 2.5 also shows the actual fraction censored overall which is $\alpha = (1 - \pi)\alpha_0 e^{-C} + \pi\alpha_1 e^{-2C}$, from where it is seen that the distribution of censoring across the covariate groups is more crucial than the overall percentage of censoring.

2.6 Discussion

The results from the simulation studies demonstrate that bootstrap methods do quite well, in terms of coverage probabilities, for estimating β^{best} in the examples considered. Particularly when applied to the Cox model estimates, both the coverage probabilities and the average interval lengths seem to be very satisfactory. The estimation procedure based on pseudo observations, however, presents difficulty in implementation and yields wider intervals. In the first comparison with multiple time

points ($k > 1$), there is perhaps an advantage to the Cox model based procedures which are being applied to a correct model. But essentially the same results are seen in the much simpler comparison with $k = 1$ where both the Cox model and the logistic model for $P_{12}(t; Z)$ are true. The results in Section 2.5 show that estimates from the PO method may not be consistent when censoring depends on the covariate. At first acquaintance, the PO method as proposed by Andersen et al. (2003) is attractive for its simplicity in implementation. However, the method should be used with great caution and particularly with censored data where modification is required if it is to give satisfactory results.

In the following, we discuss certain issues that naturally arise in the use of pseudo observations.

2.6.1 Range of the pseudo observations

It is often of interest, to study parameters that are subject to a range restriction. For example, if the parameter of interest is a probability, the parameter space is a subset of [0, 1]. However, the pseudo observations, as defined in (2.1), need not lie in this set. This phenomenon often occurs with censored data, but it can also occur with uncensored data at least in the case of the multi-state model considered here. This would not happen if $\hat{\theta}$ and $\hat{\theta}_{(-i)}$ were simple averages but the Aalen-Johansen estimator is not a simple average even in the uncensored case. See, for example, Aalen and Johansen (1978) and discussions on *embeddability* in Kalbfleisch and Lawless (1985).

The fact that the nonparametric estimators are not generally simple averages is a recurring problem in the implementation of the PO method as is discussed further below. It is seen from simulation studies that truncation of the pseudo observations introduces a bias and is not the right thing to do. However, a modification using

estimating equations also encounters difficulty with high censoring where for some samples, the ratios in (2.11) may be negative, so $\hat{\beta}$ cannot be computed. One possible remedy might be reparametrization where one considers pseudo observations corresponding to $\text{logit}\{P_{1h}(t)\}$, or any other suitable transformation that eliminates a range restriction. We have begun to look at this alternative.

2.6.2 Issues with censored data

In the context of failure models, it is natural to investigate the performance of pseudo observations for censored data and it is clear from Tables 2.2 and 2.3 and Section 2.5 that pseudo observations do not give correct answers when covariate-dependent censoring is introduced. The PO method treats the pseudo observations, whether corresponding to a censored or uncensored event, the same. In Section 2.5, an exact computation shows that the PO method is inconsistent when the censoring is not balanced in the covariate groups.

2.6.3 Modeling issues

It may not be feasible or realistic to formulate meaningful simple models for the occupancy probabilities of a multi-state model in terms of the covariates. In most models hitherto considered, these probabilities are complicated nonlinear functions of the covariates and attempts to construct a linear relationship could lead to very crude approximations. One could formulate an inverse problem of generating the data subject to the model being true (see Andrei and Murray (2007)) but this will be impossible to do for a multi-state model. This nonlinearity may partially explain why the PO method is not giving very satisfactory results, even in the complete data case. Methods using pseudo observations would work best on problems involving exact or nearly linear relationships.

An application of pseudo observations for modeling restricted mean lifetimes is suggested in Andersen et al. (2004), where a GLM is specified for the area under the survival curve, using a log link, over a specified interval, $[0, \tau]$, say. In this article, the expression for β^{best} in (2.6) is approximately proportional to the difference in the areas under the curves $P_{12}(t; Z = 1)$ and $P_{12}(t; Z = 0)$ on the logit scale, computed using equi-spaced knots. This is similar to analyzing mean lifetimes. Simulation results in Andersen et al. (2004) are limited with regard to censoring, and one should expect to see similar results with heavy or covariate-dependent censoring as in the cases considered here.

The case $k > 1$ in (2.5), even with binary Z, is only approximate and the accuracy of the approximation would depend on the choice of the time points and the shape of the probability curves (see, for example, Figure 2.2). Klein and Andersen (2005) have suggested (2.5) with $k > 1$ and different link functions but simulation studies show bias in estimates and larger mean-squared-error. Our simulation and theoretical study based on (2.8) is similar to a hypothesis testing problem for comparing two survival curves at a fixed point in time as discussed in Klein et al. (2007). Simulation results in that article show some improvement in coverage probabilities with the PO method but, even in this case, results are not satisfactory, particularly with censoring.

Finally, note that pseudo observations arise from the jackknife. Previous work in this area (see Hinkley (1977), Simonoff and Tsai (1986), Wu (1986) among many others) discusses applications where the pseudo observations are generated from the model whose parameters are to be estimated. Put simply, if estimation of a regression parameter is of interest, the regression model is re-fit each time an observation is left out and pseudo observations are defined based on this estimated regression parameter. In the applications discussed in this article, however, the pseudo obser-

vations are defined prior to fitting the regression model, substituted as the response in the regression and the regression parameter estimated. So it is not clear how one may appeal to results on the jackknife based on earlier work and more research is required in this area.

Acknowledgements

We wish to thank Hans C. Van Houwelingen, Leiden University Medical Center for useful suggestions and comments on an earlier version of this work. We are grateful to Douglas E. Schaubel for going through an earlier version of the manuscript and suggesting helpful comments and improvements. We also thank Mary K. Guidinger from Arbor Research Collaborative for Health and Randall L. Webb from The University of Michigan Kidney Epidemiology and Cost Center for helpful discussions and assistance with organizing the computational results from SRTR data.

The Scientific Registry of Transplant Recipients is funded by contract number 234-2005-37009C from the Health Resources and Services Administration, US Department of Health and Human Services. The views expressed herein are those of the authors and not necessarily those of the US Government. This study was approved by HRSA's SRTR project officer. HRSA has determined that this study satisfies the criteria for the IRB exemption described in the Public Benefit and Service Program provisions of 45 CFR 46.101(b)(5) and HRSA Circular 03.

CHAPTER III

A CUSUM to monitor outcomes of multi-center studies

3.1 Introduction

In multi-center projects involving medical interventions, it is important to monitor and provide timely outcome information to participating institutions or facilities. Such a monitoring activity can be useful in providing warning signals to the institutions and also in coordination of the project. Standard statistical techniques like average mortality, risk-adjusted mortality and multivariate modeling can be used to identify performance changes at a national level, but may be insensitive to smaller persistent changes at the facility level. The CUmulative SUM (CUSUM) procedure, introduced in Page (1954) and later discussed by Van Dobben de Bruyn (1968) and many others, is a graphical procedure for continuous sequential monitoring. It was first proposed as a technique for industrial statistical process control and its use in the context of health-care monitoring has been proposed by Novick et al. (2001), Poloniecki et al. (1998) and Woodall et al. (2006) among several others. The optimality of the CUSUM procedure in minimizing the worst possible detection delay has been discussed by Lorden (1971) and Moustakides (2004) among several others.

In the implementation of a CUSUM or other monitoring procedures, it is important to use risk-adjusted methods that can help account for institutional differences

as can arise simply due to differences in the mix of patients. Recently, Axelrod et al. (2006) and Steiner et al. (2001) have suggested the use of risk-adjusted CUSUM charts to monitor facility performance with regard to surgical interventions. In this chapter and in Axelrod et al. (2006), the methods are proposed in the context of assessing transplantation outcomes at the facility level. Although the methods apply in other contexts, we will use the transplant example in what follows.

In the above cited health-care applications of the CUSUM, the unit of analysis is a success or failure outcome associated with the transplant. Thus, each transplant is followed for a pre-specified period of time to determine whether or not a Bernoulli outcome occurs. In Axelrod et al. (2006), for example, one-year post-transplant survival status was the outcome utilized. A scoring approach based on the likelihood ratio is then used to accumulate these outcomes in a CUSUM that can be monitored for each institution. One consequence of this approach is that a death (or survival) affects the CUSUM only one year after the occurrence of the transplant. Thus, although these approaches perform well, this built-in one year lag in assessment is undesirable; a substantially more efficient CUSUM could potentially be obtained by updating the event status continuously in time. It is worth mentioning here that if the data are reported at a constant lag of one year, then the discrete time approach in Axelrod et al. (2006) may be more appropriate.

In this chapter, we present a risk-adjusted CUSUM procedure that is based on the Cox model and defined in continuous time. We shall assume, however, that reports are immediate and there is no delay involved between the occurrence of the outcome and its reporting. Simulations demonstrate that this approach tends to provide a much earlier detection of a deteriorating situation than the discrete time methods described above.

The chapter is organized as follows. Section 3.2 lays out the assumptions and framework for defining the CUSUM. Sections 3.3 and 3.4 discuss the proposed continuous CUSUM and the discrete CUSUM in Axelrod et al. (2006), respectively. Section 3.5 and 3.6 discuss several simulation studies. Section 3.7 demonstrates the CUSUM procedures on kidney transplant data from the Scientific Registry of Transplant Recipients (SRTR). The chapter concludes with a short discussion in Section 3.8.

3.2 Notation and assumptions

Consider a specific institution or facility and suppose that transplants occur over time beginning at some time origin denoted by $t = 0$. Let the successive times of transplant be $S_1, S_2, \ldots$, for patients numbered $1, 2, \ldots$, and let $N^A(t)$ denote the number of transplants in $(0, t]$. We will later model $\{N^A(t), t \geq 0\}$ as a Poisson process, but for the moment we work conditionally on the times. In what follows, we measure time in years.

Let X_i denote the time from transplant to graft failure or death, and assume that X_i and S_i are independent given covariates Z_i that are measured at the time of transplant. The chronological time of failure is $T_i = S_i + X_i$. We suppose that, conditional on the covariates Z_i, there is a known null distribution of interest for X_i which is defined by the hazard function $\alpha_i(x)$. In the applications we consider, this distribution is estimated from the experience of all transplant centers combined. We have used a Cox regression model based on the measured covariates Z_i, which means $\alpha_i(x) = \lambda_0(x)e^{Z_i^T\beta}$ for $x > 0$. Thus, $\alpha_i(x)$ represents a national average failure rate for an individual with covariate Z_i. Other regression models could, of course, be used. In what follows, we assume that the estimates obtained for α_i are without

error; in fact, the estimation errors will be very small when the estimates arise from the very large data set of all transplant recipients nationwide. In other applications, one may want to take account of the uncertainty in the estimation of α_i.

Let $\tilde{N}_i^D(t) = I(T_i \leq t)$ be the failure time counting process for the i-th transplant where t is chronological time. We also let $Y_i(t)$ be an at-risk indicator, which determines whether or not the i-th transplant could give rise to a qualifying event. By a qualifying event, we shall mean a transplant failure that occurs within one year from the date of transplant. Since time is in years, this means, $Y_i(t) = I(S_i < t \leq (S_i + 1) \wedge T_i)$. We also define $N_i^D(t) = \int_0^t Y_i(u) d\tilde{N}_i^D(u)$ for $t > 0$. Note that $N_i^D(t)$ is 0 for all t unless a transplant failure occurs within one year of the transplant date, in which case it jumps to 1 at the (chronological) time of failure. We define the history (or filtration) for $N_i^D(t)$ as $\mathcal{F}_{t-} = \sigma\{N^A(u), N_i^D(u), Y_i(u), Z_i, N^A(t), i = 1, 2, \ldots, n_t = N^A(t) : 0 \leq u < t\}$. If the hypothesized null rate holds,

$$\begin{aligned} E(dN_i^D(t)|\mathcal{F}_{t-}) &= d\Lambda_i(t) \\ &= Y_i(t)\alpha_i(t - S_i)dt \end{aligned} \tag{3.1}$$

Clearly $\mathcal{F}_{t-}$ specifies S_i for all $i \leq n_t$.

For the institution under study, let $N^D(t) = \sum_{i \geq 1} N_i^D(t)$ be the observed number of qualifying transplant failures that have occurred up to time t.

3.3 A CUSUM in continuous time

As is usual, the CUSUM we define is based on the Sequential Probability Ratio Test (SPRT) introduced by Wald (1947). We consider the process $N^D(t),\ t > 0$ as the response so that the likelihood contribution in any small interval $(t, t + dt]$ takes

into account the information about all transplants at risk of a qualifying failure at time t.

3.3.1 Definition

As above, suppose we have risk-adjusted population values for the hazard $\alpha_i(t)$ for each transplant i at risk of a qualifying failure at chronological time $t > 0$. For a selected constant $\theta > 0$, we consider a likelihood ratio statistic corresponding to a test of $\theta = 0$ versus $\theta > 0$. Thus the null intensity function for the ith subject is $E[dN_i^D(t)|\mathcal{F}_{t-}] = d\Lambda_i(t)$ as in (3.1). The likelihood ratio, based on data $dN_i^D(t)|\mathcal{F}_{t-}$, with intensity $e^\theta d\Lambda_i(t)$ versus $d\Lambda_i(t)$ is

$$\prod_{i\geq 1}\left[\frac{e^{\theta dN_i^D(t)}[1-e^\theta d\Lambda_i(t)]^{1-dN_i^D(t)}}{[1-d\Lambda_i(t)]^{1-dN_i^D(t)}}\right]$$

where we have used (3.1) and a binomial likelihood for the differential increment $dN_i^D(t)$. By a repeated conditioning argument, the full likelihood upto time t is a limit, in the product integral sense, of the likelihoods in $(t, t+dt]$. We therefore obtain, after taking the logarithm, an expression for the CUSUM increment at time t as,

$$dU_t \;=\; \theta dN^D(t) - (e^\theta - 1)dA(t) \tag{3.2}$$

where we define $dA(t) = \sum_{i\geq 1} d\Lambda_i(t)$. If a two-sided alternative was of interest, we would compute the CUSUM as $U_t = \int_0^t dU_s$ and strong trends, either up or down, would provide a signal. In the applications of interest here, however, only the one-sided alternative in which the CUSUM trends upward is of concern. This is a continuous time extension of the usual discrete time CUSUM introduced in Page (1954) and utilized in Axelrod et al. (2006) and Steiner et al. (2001), for example. The CUSUM suitable for the one-sided alternative is $\{G_t\}$ where $G_0 = 0$ and

$$G_{t+dt} \;=\; \max(0, G_t + dU_t), \qquad \text{for } t > 0.$$

It can be seen that this is equivalent to

$$G_t \;=\; U_t - \min_{0\le s\le t} U_s, \quad t>0$$

where U_t is

$$U_t \;=\; \theta N^D(t) - (e^\theta - 1)\int_0^t dA(u), \quad t>0 \tag{3.3}$$

It is easy to see that U_t has positive jumps only when there is a qualifying failure, otherwise it trends downwards. Note also that $\min_{0\le s\le t} U_s$ is non-increasing and changes in value only when $G_t = 0$. It is of some interest to note that (3.2) is exactly equal to the log-likelihood ratio increment for a testing problem in which $N^D(t)$ is a Poisson process with mean function $A(t) = \int_0^t dA(u)$.

To implement the CUSUM procedure, we monitor G_t in continuous time until it crosses a fixed upper barrier at $h > 0$, at which time a signal is given. The value h is chosen so as to give some suitable properties to the monitoring scheme. Let the time until the barrier at h is reached be

$$\tau_h \;=\; \inf\{t>0 : G_t \ge h\}.$$

In the CUSUM literature, $E(\tau_h)$ is referred to as the Average Run Length (ARL) and plays an important role in designing parameter values for the CUSUM. In the next section, we shall discuss a theoretical approximation to $E(\tau_h)$ under certain assumptions about the arrival process $N^A(t)$.

The above CUSUM has been obtained by considering a sequential probability ratio test that the intensity for $dN_i^D(t)$, conditional on past history, is $e^\theta d\Lambda_i(t)$ versus $d\Lambda_i(t)$. In the next subsection, we consider the properties of this CUSUM under the assumption that the true rate for the facility under study is

$$E(dN_i^D(t)|\mathcal{F}_{t-}) \;=\; e^\mu d\Lambda_i(t), \tag{3.4}$$

which means that the facility of interest has a rate of failure that is e^{μ} times the national or population mean rate.

3.3.2 An approximation for U_t

For the purpose of developing an approximation, we start with (3.3). It is natural to assume that new transplants occur according to a homogeneous Poisson process with rate ψ say, and in order to study the process in equilibrium, we suppose that the arrivals begin at time $t = -1$. This is sufficient for equilibrium since only failures within one year of transplant are viewed as qualifying events. Again we denote the transplant times as $S_1, S_2, \ldots$, but since we begin at time $t = -1$, S_i+1 has a Gamma distribution with scale ψ and shape i. We assume that the sequence of covariates, $Z_1, Z_2, \ldots$, are independent and identically distributed (iid) and that the hazard for X_i conditional on Z_i is $e^{\mu}\alpha_i(x)$ for $x > 0$ which leads to (3.4). We find that

$$\begin{aligned}
E[dA(u)] &= E\left[\sum_{i\geq 1} Y_i(u)\alpha_i(u - S_i)du\right] \\
&= e^{-\mu}\sum_{i\geq 1} E\left[I(u \leq S_i + 1 < u + 1, X_i \geq u + 1 - (S_i + 1))\right. \\
&\qquad \left. \times e^{\mu}\alpha_i(u + 1 - (S_i + 1))\right] du \\
&= e^{-\mu}\sum_{i\geq 1}\int_u^{u+1} E\{f_\mu(u + 1 - x|Z_i)\}\psi\frac{e^{-\psi x}[\psi x]^{i-1}}{(i-1)!}dx\, du \\
&= e^{-\mu}\int_u^{u+1} E\{f_\mu(u + 1 - x|Z_i)\}\psi\sum_{i\geq 1}\frac{e^{-\psi x}[\psi x]^{i-1}}{(i-1)!}dx\, du \\
&= e^{-\mu}\psi E[F_\mu(1|Z_i)]\, du \qquad (3.5)
\end{aligned}$$

Here $f_\mu(\cdot|Z_i), F_\mu(\cdot|Z_i)$ are the pdf and cdf of $X_i|Z_i$, respectively, and the final expectation is with respect to the marginal distribution of Z_i at the facility under study. Thus $E[dA(u)] = \gamma\, du$, where $\gamma = e^{-\mu}\psi E[F_\mu(1|Z_i)] > 0$ is a constant.

On the other hand, it is easily seen that $N^D(t)$ is a homogeneous Poisson process

with mean $\psi E[F_\mu(1|Z_i)]$. In summary, we have the following result,

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

where $N^D(t),\ t > 0$ is a homogeneous Poisson process with rate $e^\mu\gamma = \psi E[F_\mu(1|Z_i)]$ and $E_t = (e^\theta - 1)[\gamma t - A(t)]$ is a zero-mean process. In order to get a theoretical approximation to $E(\tau_h)$ in the next section, we make the approximation $E_t = 0$ to obtain theoretical average run lengths and later assess these approximations in simulations.

3.3.3 An approximation to the average run length

The CUSUM process G_t may be thought of as the process U_t with barriers at 0 and $h > 0$. The barrier at h is absorbing, but the barrier at 0 needs special attention since we do not terminate the CUSUM when it reaches 0. When G_t becomes 0 it stays at 0 until U_t registers a jump of size θ, at which time G_t jumps up by θ as well. Since U_t has stationary increments, the process G_t can be thought of as undergoing a renewal, each time this jump from 0 to θ occurs.

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

$$F_0 \;=\; \inf\{t > 0 : G_t = \theta\}. \tag{3.6}$$

F_0 is the time of the first event in the Poisson process $N^D(t)$ and so has an Exponential distribution with rate $e^\mu\gamma$. Thus, we can think of the sequence of events $[G_t = \theta, G_{t-} = 0]$ as constituting a renewal process delayed by F_0. Following each renewal there are two possibilities: either there is another return to zero (and a subsequent renewal) or the process terminates by crossing the barrier $h > 0$.

First, consider the process U_t with initial value $U_0 = \theta$ and with absorbing barriers at 0 and h. Let p_R and $(1 - p_R)$ be the respective probabilities of absorption at h

and 0. Further, let

$$T^{(\theta)} = \inf\{t > 0 : U_t \notin (0, h), U_0 = \theta\} \tag{3.7}$$

be the time to absorption.

In the process G_t, let J represent the number of renewals (i.e., occurrences of $[G_t = \theta, G_{t-} = 0]$) including the first so that $J \geq 1$ and note that J is a stopping time. It is easy to see that

$$P(J = j) = (1 - p_R)^{j-1} p_R, \qquad j \geq 1.$$

Following 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, and $T_i^{(\theta)}$ is the time until the process exceeds h or returns to 0, whichever occurs first. It is now easy to see that $E(W_i) = E(T^{(\theta)}) + (1 - p_R)E(F_0)$ since $E(R_i) = E(F_0)$. Thus, $\tau_h = F_0 + \sum_{i=1}^{J} W_i$ and, since J is a stopping time, an application of Wald's identity gives

$$\begin{aligned} E(\tau_h) &= E(F_0) + E(J)[E(T^{(\theta)}) + (1 - p_R)E(F_0)] \\ &= E(F_0) + \frac{E(T^{(\theta)}) + (1 - p_R)E(F_0)}{p_R} \\ &= \frac{E(T^{(\theta)}) + E(F_0)}{p_R} \end{aligned} \tag{3.8}$$

Since $E(F_0) = e^{-\mu}/\gamma$, it only remains to find p_R and $E(T^{(\theta)})$.

Consider again the process $\{U_t\}$ with $U_0 = 0$ and with absorbing barriers at $-\theta$ and $h - \theta$. Let $f^*(\omega) = E(e^{-\omega U_t})$ be the moment-generating function of U_t in the unrestricted process, so that

$$f^*(\omega) = \exp\{\gamma t[\omega(e^{\theta} - 1) + e^{\mu}(e^{-\omega\theta} - 1)]\}.$$

Let $\omega_0 \neq 0$ satisfy $f^*(\omega_0) = 1$ which, it should be noted, does not depend on t. From Wald's identity for restricted random walks in continuous time (see, for example, Cox and Miller (1965)), we have, for any ω,

$$E \exp\{-\omega U_{T^{(\theta)}} - T^{(\theta)} \log f^*(\omega)\} = 1$$

and ignoring overshoot across the boundaries, this gives

$$e^{-\omega_0(h-\theta)} p_R + (1 - p_R) e^{\omega_0 \theta} \approx 1$$

so that

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

Let $\eta\, dt = E(dU_t)$, which may be interpreted as the drift parameter for this walk. It is easy to see that

$$\eta = (\theta e^{\mu} - e^{\theta} + 1)\gamma. \tag{3.9}$$

If $\eta \neq 0$, then Wald's identity gives, $\eta\, E(T^{(\theta)}) = E(U_{T^{(\theta)}})$ where, again ignoring overshoot,

$$E(U_{T^{(\theta)}}) \approx (h - \theta) p_R + (-\theta)(1 - p_R) = h p_R - \theta, \ \eta \neq 0.$$

When $|\eta| \to 0$, then $|\omega_0| \to 0$ and $p_R \to \theta/h$. If $\sigma^2 dt = \mathrm{Var}(dU_t)$, then (with $\eta = 0$) Wald's identity gives

$$E(U^2_{T^{(\theta)}}) = \sigma^2 E(T^{(\theta)})$$

and we have $E(U^2_{T^{(\theta)}}) \approx (h - \theta)^2 p_R + \theta^2 (1 - p_R)$. Noting that $\sigma^2 = \theta^2 \gamma e^{\mu}$ and after some simplification, we obtain the approximations

$$E(\tau_h) \approx \begin{cases} \frac{h}{\eta} - \frac{e^{-\mu}(e^{\theta} - 1)}{\eta} \left(\frac{1 - e^{-\omega_0 h}}{1 - e^{-\omega_0 \theta}} \right), & \eta \neq 0 \\ \\ \frac{h^2 e^{-\mu}}{\theta^2 \gamma}, & \eta = 0. \end{cases} \tag{3.10}$$

It can be shown that the limit of $E(\tau_h)|_{\eta\neq 0}$ as $\eta \to 0$ is $E(\tau_h)|_{\eta=0}$ so that $E(\tau_h)$ is continuous in η. The arguments are summarized in Appendix B.

3.4 A CUSUM in discrete time

In this section, we consider CUSUM procedures in discrete time that include those discussed in Axelrod et al. (2006) and Steiner et al. (2001). These CUSUMs are also based on the SPRT, but are defined with reference to the transplant or arrival times $\mathcal{S} = \{0 = S_0, S_1, S_2, \ldots, \}$ of the process $N^A(t)$. Each transplant gives rise to a binary experiment resulting in a success or a failure depending on whether or not the transplant survives for a period of one year. We define the process on the integers where i corresponds to the i-th transplant, $i = 1, 2, \ldots$. In defining the continuous time CUSUM in Section 3.3, we considered a likelihood ratio statistic based on a hypothesized relative risk or hazard ratio, that related the failure rates of the institution of interest to those of the population as a whole. A similar approach can be used in discrete time where we consider a formulation using an odds ratios as discussed in Axelrod et al. (2006) and Steiner et al. (2001) or a discrete relative risk.

We use a notation similar to that in Section 3.2 and define

$$\xi_i \;=\; I(X_i \leq 1), \qquad i = 1, 2, \ldots$$

Let $\pi_{i0} = E(\xi_i = 1|Z_i)$ be the population probability of failure in one year at the given covariate Z_i, and suppose that these are known or accurately estimated from a large sample. As before, we assume that the covariates $Z_1, Z_2, \ldots$ in the institution of interest are iid.

3.4.1 Testing the odds ratio

First, we consider an odds ratio model

$$\text{logit } \pi_{i1} = \text{logit } \pi_{i0} + \log OR \tag{3.11}$$

in which the odds ratio OR measures the difference between the institution of interest and the overall national rates. Consider a likelihood ratio test of $H_0 : OR = 1$ versus $H_1 : OR = OR_A > 1$. The contribution to the log-likelihood ratio from ξ_i is,

$$\begin{aligned} g_i &= \log\left[\left(\frac{\pi_{i1}}{\pi_{i0}}\right)^{\xi_i}\left(\frac{1-\pi_{i1}}{1-\pi_{i0}}\right)^{1-\xi_i}\right] \\ &= \xi_i(a_i + b_i) - b_i \end{aligned}$$

where $a_i = \log(OR_A) - \log(1 - \pi_{i0} + \pi_{i0}OR_A)$ and $b_i = \log(1 - \pi_{i0} + \pi_{i0}OR_A)$. The corresponding CUSUM is defined by $G_0 = 0$ and $G_{i+1} = \max(0, G_i + g_i),\ i = 1, 2, \ldots$. For h chosen to obtain desirable properties, the procedure generates a signal at 'time' κ_h where

$$\kappa_h = \min\{k \geq 1 : G_k \geq h\}.$$

Here, the ARL is defined as $E(\kappa_h)$ which is the average number of transplants carried out and assessed before a signal is obtained.

We examine properties of the CUSUM for the facility of interest under the assumption that the true odds ratio is $OR = e^{\mu}$. For this case, we have independent realizations of $\{\xi_i\}_{i\geq 1}$ with mean $p_\mu = E(\xi_i | OR = e^{\mu})$.

Since $a_i + b_i = \log(OR_A)$, g_i has the same structure as the differential increment dU_t in (3.2) for the continuous time CUSUM. The Bernoulli variable ξ_i is the discrete analog of $dN^D(t)$ and $\log(OR_A)$ plays the same role as the log relative risk θ in (3.2). We find that $\zeta = E(g_i) = E[\xi_i \log(OR_A) - b_i] = p_\mu \log(OR_A) - E(b_i)$. For deriving a theoretical approximation, we will work with $\tilde{g}_i = \xi_i \log(OR_A) - E(b_i)$, arguing exactly as for the continuous case.

3.4.2 Testing the relative risk

Consider now, the alternative discrete relative risk (RR) model

$$\log \pi_{i1} = \log RR + \log \pi_{i0} \tag{3.12}$$

and a test of the hypothesis $H_0 : RR = 1$ versus $H_1 : RR = RR_A > 1$. The contribution from ξ_i to the log-likelihood ratio statistic is,

$$\begin{aligned} g_i &= \log\left[\left(\frac{\pi_{i1}}{\pi_{i0}}\right)^{\xi_i}\left(\frac{1-\pi_{i1}}{1-\pi_{i0}}\right)^{1-\xi_i}\right] \\ &= \xi_i(a_i+b_i) - b_i \end{aligned}$$

where $a_i = a = \log(RR_A)$ and $b_i = \log(1-\pi_{i0}) - \log(1-RR_A\pi_{i0})$. Here again, we consider independent realizations of $\{\xi_i\}_{i\geq 1}$ under the assumption that the true relative risk for the institution of interest is $RR = e^{\mu}$. We find that $\zeta = E(g_i) = E[(a_i+b_i)F_{\mu}(1|Z_i) - b_i]$ with the expectation being taken with respect to the common distribution of the Z_i. Theoretical approximations are based on step sizes $\tilde{g}_i = a\xi_i + [1-\xi_i]E(b_i)$.

3.4.3 Approximating the average run length

The process $\{V_k = \sum_{i=1}^{k} g_i\}_{k\geq 1}$ is a random walk and we impose barriers at 0 and $h > 0$. At the i-th step, the walk makes a jump of $g_i = \xi_i(a_i+b_i) - b_i = \xi_i(a+b) - b + e_i$. Here $a = E(a_i), b = E(b_i)$ are constants. For the purpose of deriving a theoretical approximation to $E(\kappa_h)$, we shall assume $e_i = 0$ and compare this approximation in simulations. For the CUSUM based on (3.11), $e_i = b - b_i$ and has mean zero. Also, h is an absorbing barrier and when V_k reaches 0, it stays at 0 for some time until a qualifying failure is observed. We give an approximation to $E(\kappa_h)$ by applying the same technique as in Section 3.3.3.

The moment-generating function of the step size g_i is,

$$f^*(\omega) \;=\; p_\mu e^{-\omega a} + (1-p_\mu)e^{\omega b}$$

where the right-hand side is obtained after taking an expectation with respect to Z_i and hence is free of i. We let $\omega_0 \neq 0$ be such that $f^*(\omega_0) = 1$. The mean step size is $\zeta = E(g_i)$ and we also let $\sigma_d^2 = \mathrm{Var}(g_i)$. It is easy to see that $\zeta = p_\mu(a+b) - b$, and $\sigma_d^2 = (a+b)^2 p_\mu(1-p_\mu)$. We define F_0^d as,

$$F_0^d = \min\{k \geq 0 : \xi_{k+1} = 1\}$$

which is the waiting time to the first occurrence of $\xi_i = 1$ and therefore has a Geometric distribution with success probability p_μ. We define $T^{(a)}$ as,

$$T^{(a)} \;=\; \min\{k \geq 1 : V_k \notin (0,h), V_1 = a\}$$

which is the time to absorption at either 0 or h, starting at a. Note that F_0^d, $T^{(a)}$ are the discrete analogs of (3.6) and (3.7) and a relationship similar to (3.8) holds for $E(\kappa_h), E(F_0^d)$ and $E(T^{(a)})$. This yields the following approximation,

$$E(\kappa_h) \;\approx\; \begin{cases} \frac{h}{\zeta} - \frac{b(1-p_\mu)}{\zeta p_\mu}\left(\frac{1-e^{-\omega_0 h}}{1-e^{-\omega_0 a}}\right), & \zeta \neq 0 \\ \\ \frac{h^2}{ab} + \frac{h}{a}, & \zeta = 0 \end{cases} \tag{3.13}$$

where, we have used $p_\mu = b/(a+b)$ and $\sigma_d^2 = ab$ for $\zeta = 0$. We define

$$\tau_h^d \;=\; 1 + \sum_{i=1}^{\kappa_h}(S_i - S_{i-1})$$

as the chronological time at which the discrete time CUSUM signals. Observe that we add 1 to the expression to account for the inherent 1 year lag in the discrete approach. This helps to compare the discrete and continuous time CUSUMs. Under

the Poisson model, $\{S_i - S_{i-1}\}_{i\geq 1}$ are iid Exponential random variables with mean $1/\psi$ and so,

$$E(\tau_h^d) \;=\; 1 + \frac{E(\kappa_h)}{\psi}$$

3.5 Simulation study

In the simulations, we considered a facility where transplants occurred according to a homogeneous Poisson process at a rate of $\psi = 100$ transplants per year. We supposed that the the post-transplant failure time distribution at the nationwide or population level was Exponential with rate λ_0 and chose a one-year failure probability of

$$1 - e^{-\lambda_0} = 10\%, \tag{3.14}$$

in approximate agreement with the national data as summarized in the next section. In assessing the procedures, we supposed that the post-transplant failure time X_i at the specific facility under review had a hazard of $e^{\mu}\lambda_0$. We compared the performance of the CUSUM procedures discussed above in terms of average run length (ARL), and also assessed the accuracy of the approximations obtained.

In the continuous CUSUM of Section 3.3, we chose $\theta = \log 2.0$. Similarly, in the discrete time CUSUMs defined in (3.11) and (3.12), we chose $OR_A = 2.0$ and $RR_A = 2.0$, respectively, thereby ensuring that all the three CUSUM procedures are testing similar hypotheses. By the design of the discrete time CUSUM, the outcome (failure or success) of a transplant is not counted until a year is past from the time of transplant. In reality, therefore, the CUSUM is evaluated and capable of sending a signal at times $\{S_1 + 1, S_2 + 1, \ldots\}$. To reflect this, 1 year was added to the average run lengths based on the times $\{S_1, S_2, \ldots\}$. Accordingly, the theoretical

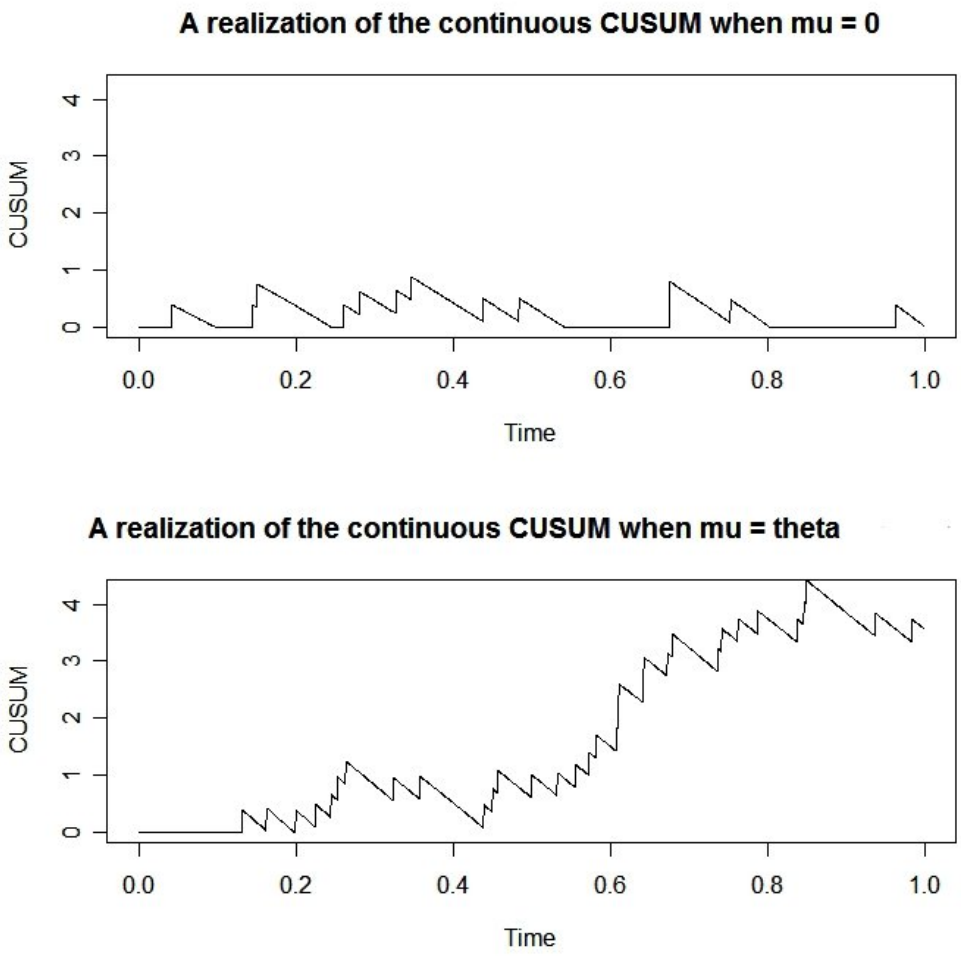

Figure 3.1: Realizations of the continuous CUSUM under intensities $\alpha_i(\cdot)$ and $e^{\theta}\alpha_i(\cdot)$.

approximations in (3.13) were also increased by one.

Sample realizations of the continuous time CUSUM with $\mu = 0$ and $\mu = \theta$ in (3.4) are shown in Figure 3.1.

In the simulations, all the CUSUMs were followed until a signal (i.e., a crossing of the control limit, $h > 0$) was reached. The choice of the control limit for each of the three CUSUMs was done in a way so as to ensure that all of them have approximately the same ARL when $\mu = 0$. While one may also use other criteria to choose control limits for CUSUM charts, this particular criterion is essential in order to make a meaningful comparison of the ARL across the different CUSUM charts.

Table 3.1: ARLs for the three CUSUM procedures for different choices of facility relative risk e^{μ}.

e^{μ}	η	Average τ_h	Median τ_h	$E(\tau_h)$	$\widehat{SD}(\tau_h)$
Continuous-time CUSUM $(h = 4.35, e^{\theta} = 2.0)$					
0.69	-4.40	518.51	428.52	410.32	394.26
0.82	-3.36	142.00	91.62	105.12	143.94
1.00	-1.89	29.92	20.92	23.62	29.13
1.22	-0.09	8.07	5.76	6.59	7.66
1.50	2.16	2.72	2.08	2.43	2.34
1.73	4.05	1.50	1.22	1.45	1.09
2.00	6.21	0.98	0.83	0.97	0.62
Discrete-time CUSUM, OR $(h = 4.19, OR_A = 2.0)$					
0.69	-4.40	469.09	347.08	499.45	383.53
0.82	-3.36	146.76	102.00	119.40	142.70
1.00	-1.89	29.75	21.38	26.24	27.79
1.22	-0.09	8.63	6.20	7.62	7.23
1.50	2.16	3.75	3.15	3.43	2.19
1.73	4.05	2.48	2.19	2.46	1.06
2.00	6.21	1.97	1.86	1.98	0.58
Discrete-time CUSUM, RR $(h = 4.47, RR_A = 2.0)$					
0.69	-4.40	466.53	317.22	372.19	469.81
0.82	-3.36	119.19	85.24	107.66	114.64
1.00	-1.89	29.64	22.18	24.83	27.44
1.22	-0.09	8.86	6.71	7.83	7.41
1.50	2.16	3.85	3.06	3.51	2.49
1.73	4.05	2.52	2.19	2.47	1.17
2.00	6.21	2.02	1.84	1.97	0.70

Thus, when a facility is performing at the national average we chose h to achieve an ARL of around 30 years, which corresponded to $h = 4.35,\ h = 4.19$, and $h = 4.47$ for the continuous, discrete logistic and discrete relative risk CUSUMs, respectively. We shall see later that this choice for the ARL also results in some interesting properties for the continuous CUSUM.

Table 3.1 shows the results from 1000 repetitions at various values of μ and for each of the three CUSUMs. The third, fourth and sixth columns give the sample average, median and standard deviation based on the 1000 replications. The fifth column gives the approximation to the ARL based on the appropriate formulas in

(3.10) and (3.13). Generally speaking, the theoretical approximation seems to work well when the random walk process U_t has a positive drift ($\eta > 0$).

In comparison to the discrete time CUSUMs, the continuous time CUSUM gives shorter run lengths or signal times when $e^{\mu} > 1$ (facility is doing worse than the national rate) and longer run lengths when $e^{\mu} < 1$. The two discrete time CUSUMs perform similarly for higher values of the relative risk or odds ratio. The unit of time in all these simulations is years, so an average run length of 30 years means that one would expect to wait 30 years on average to get a signal. Since we are using an average rate of 100 transplants per year, this corresponds to an average of about 3000 transplants to the first signal under the null rate of $\mu = 0$. It may be noted here that the number of transplants until signal possesses an invariance property, which simply means that if the rate of transplants ψ is halved to 50 per year, then the ARL, as obtained from simulations and (3.10), would get doubled thereby keeping the number of transplants until signal as constant. Thus, for purposes of selecting h, the number of transplants until signal may be a better measure.

Since we start monitoring all the CUSUMs at the same time and, for the discrete CUSUMs, follow the transplants a year into the future, the comparison between the continuous and discrete methodologies seems fair in the sense that all of them use the same amount of data. Both the continuous and discrete CUSUMs account for failures within the first year of observation; the gain arises, however, since the continuous CUSUM counts the failure at the time it occurs whereas the discrete CUSUM responds to the failure only a year after the transplant date.

A study of how empirical ARLs compare with theoretical approximations in (3.10) and (3.13) across different choices of the control limit h and the parameter μ was also done. We report the results for the continuous time CUSUM and the discrete (OR)

Table 3.2: Empirical and theoretical ARLs for discrete (OR) and continuous CUSUMs for different choices of control limit h and facility relative risk e^{μ}.

e^{μ}	$h = 3.0$		$h = 4.0$		$h = 5.0$		$h = 6.0$	
	ARL	$E(\tau_h)$	ARL	$E(\tau_h)$	ARL	$E(\tau_h)$	ARL	$E(\tau_h)$
				Continuous time CUSUM				
1.00	19.07	17.16	59.21	52.69	161.60	151.12	407.81	420.48
1.22	5.21	4.81	9.08	8.63	13.93	13.65	20.63	19.93
1.50	2.11	2.02	3.03	2.94	4.04	3.88	4.79	4.82
1.73	1.31	1.31	1.81	1.82	2.33	2.34	2.93	2.85
2.00	0.93	0.94	1.25	1.28	1.61	1.62	1.97	1.96
				Discrete time CUSUM, OR				
1.00	24.74	21.18	67.40	62.86	174.36	178.25	477.63	493.99
1.22	6.81	6.23	10.74	10.14	16.04	15.11	21.91	21.12
1.50	3.15	3.15	4.08	4.23	4.93	5.03	6.00	5.99
1.73	2.41	2.39	2.91	2.92	3.50	3.45	3.96	3.98
2.00	1.98	2.00	2.34	2.35	2.68	2.71	3.12	3.06

CUSUM of (3.11) in Table 3.2. From this, one can see that the accuracy of the theoretical approximations increase with μ and h. In the case of the continuous CUSUM, the percentage of discrepancy between the empirical and theoretical estimates ranges from around 11% for $h = 3.0$ to about 3% for $h = 6.0$ (for $\mu = 0.0$). This may be attributed to the fact that the error incurred in ignoring the overshoot across the boundaries for Wald's approximations decreases with increasing h. Therefore, it may be noted that if one wished to design the CUSUMs to achieve certain ARLs for values of $e^{\mu} > 1$, the formulas in (3.10) and (3.13) are accurate and could be used directly to obtain a suitable h. The values of the ARL reported so far also depend crucially on λ_0 which is chosen in (3.14) in accordance with an overall national failure rate of 10%.

So far, for the simulation study, we have assumed $\beta = 0$ and also that the covariate distribution at the facility is the same as in the population. We implemented a simulation study assuming $\beta = 0.2$ and $Z_i \overset{iid}{\sim} N(0, 1)$ in the population. λ_0 was

Table 3.3: Table showing variation in ARLs for the continuous CUSUM across facility covariate distribution parameters at $h = 2.30$.

			e^{μ}				
(ν, ρ^2)	$E_{\nu,\rho}[F_{\mu=0}(1\|Z_i)]$	$F_{\nu,\rho} \times 0.10$	1.00	1.22	1.50	1.73	2.00
(0, 1)	0.100	0.100	8.21	3.04	1.44	0.98	0.69
(0.5, 1)	0.110	0.110	7.80	2.73	1.27	0.90	0.65
(1, 1)	0.120	0.122	7.22	2.50	1.22	0.80	0.58
(0, 0.25)	0.098	0.098	8.53	3.01	1.47	0.96	0.72
(0.5, 0.25)	0.108	0.109	7.58	2.81	1.35	0.89	0.64
(1, 0.25)	0.119	0.120	7.15	2.71	1.21	0.78	0.59

chosen to satisfy

$$E[1 - e^{-\lambda_0 e^{\beta Z_i}}] \;=\; 10\%, \tag{3.15}$$

a condition similar to (3.14). At the facility of interest, we assumed $Z_i \overset{iid}{\sim} N(\nu, \rho^2)$ and computed empirical ARLs for various μ with the continuous CUSUM at $h = 2.30$. Table 3.3 shows the results. The second column shows $E_{\nu,\rho}[F_{\mu}(1|Z_i)]$, computed with respect to the facility covariate distribution. It can be shown that $E_{\nu,\rho}[F_{\mu}(1|Z_i)] \approx F_{\nu,\rho} \times E_{0,1}[F_{\mu}(1|Z_i)]$, with $F_{\nu,\rho} = \exp\{\beta\nu + \frac{1}{2}\beta^2(\rho^2 - 1)\}$. Using (3.15) we get, for $\mu = 0$, $E_{\nu,\rho}[F_{\mu}(1|Z_i)] \approx F_{\nu,\rho} \times 0.10$. The third column shows this approximation, for $\mu = 0$, and it seems to be working well. The first row in Table 3.3 represents the case when the covariates at the facility have the same distribution as the population. The next few rows show alterations to the empirical ARLs as the location and scale parameters change for the facility covariate distribution. It is evident that the effect of location ν may be substantial compared to the scale ρ in terms of modifying the ARLs. It is also worth noting here that $F_{\mu}(1|Z_i)$ is the one-year failure probability, so shifts in the covariate distribution are to be assessed only up to a shift in this 1 year failure probability. The constant $F_{\nu,\rho}$ may be seen as a quick estimate for $E_{\nu,\rho}[F_{\mu}(1|Z_i)]$ which is to be used in (3.5) and henceforth in the formulas (3.10). Also, $1/F_{\nu,\rho}$ may be interpreted as the approximate factor by which the theoretical

ARLs (3.10) at the facility would shrink or dilate compared to the population with the bonus being that this factor is free of μ, the facility outcome effect.

3.6 Controlling the rate of false alarms

Until now, we have discussed the ARL only as a performance measure and using the in-control or null ARL to set a control limit. In this section we shall discuss false alarm rates for CUSUM procedures and assess the performance of the aforementioned CUSUM charts.

In- and out-of-control states may be quantified as the mean of the observations or more specifically as the number of standard deviations from a fixed mean. For our present problem, we consider $\mu = 0$ as corresponding to an in-control state and $\mu = \theta$ to an out-of-control state. Similarly, for the discrete CUSUMs, $OR[RR] = 1$ and $OR[RR] = OR_A[RR_A]$ correspond to in- and out-of-control states, respectively. In designing a CUSUM chart for a particular application, a crucial parameter to be chosen is the control limit h for the chart. Choosing a limit very low would lead the CUSUM to signal more often, thus increasing the chance of raising false alarms. Setting it too high would lead to very long waiting times for a signal, thereby increasing the chances of late detection of an out-of-control state. Therefore, the choice of a control limit should be motivated by a tradeoff between the percentage of false positives and that of true positives. It is important that the rate of false positives be kept low and that of true detection kept high, bearing in mind the objective of quality improvement.

The concepts are similar to the notion of Type-I error and power in traditional hypothesis testing for fixed sample sizes. However, for statistical process control techniques like the CUSUM, the Type-I error is a function of time (or equivalently,

sample size) in that it measures the chance of obtaining a signal by a fixed time. As time increases, this will approach 1 as the CUSUM chart will eventually cross any fixed boundary. For this reason, it is useful to consider instead, quantities like the in-control ARL and the out-of-control ARL. When the process is in control, the CUSUM should be designed to have a long ARL or in-control ARL and when out-of-control, a short ARL or out-of-control ARL.

In the following, we shall consider a single CUSUM chart, and discuss simulation results showing the variation of in- and out-of-control ARLs with the control limit as well as estimates of the false alarm rate. As the discrete (RR) CUSUM is almost similar in performance to the discrete (OR) CUSUM, we shall study the continuous and the discrete (OR) CUSUM charts only.

Recall that the underlying test of hypothesis is that of $H_0 : \mu = 0$ vs. $H_1 : \mu = \theta > 0$. Denote the control limit by h and define ARL_0^h as the in-control ARL (under H_0) and ARL_1^h as the out-of-control ARL (under H_1). The following discussion pertains to a method adopted in Marshall et al. (2004). Consider a fixed decision time interval $[0, T]$ and define the Type-I error by time T as the proportion of false alarms under H_0 by time T. This translates to the proportion of times that $ARL_0^h < T$. Define also, the power by time T, as the proportion of successful alarms/detections under H_1 by time T which would be the proportion of times that $ARL_1^h < T$. Clearly, if we want to set a control limit h based on restricting the Type-I error to be no more than α, where $0 < \alpha < 1$ is fixed, then we need to compute the distribution of ARL_0^h. One way to do this is to choose h such that the lower α percentile of the distribution of ARL_0^h is at most T.

In Table 3.4 we present results from a simulation study that sets control limits for the two CUSUMs using the above procedure with $T = 5$ years and allowing a

Type-I error of around $\alpha = 15\%$. The choice of $\alpha = 15\%$ arises from consideration of procedures currently used by SRTR for signaling centers (see Dickinson et al. (2006)) where essentially a significance test at level 5% is conducted every 6 months for moving cohorts of size 2.5 years. Therefore, for $T = 5$, the chance of getting a signal under the null could be around 10-15%. Clearly, the results would depend on the number of transplants that have occurred by this time, or equivalently the rate of transplants ψ per year. We therefore considered $\psi = 10, 25, 40, 60, 80$ and 100, and identical assumptions on the transplants and failure time distribution as in Section 3.5. We considered the alternative again as $e^{\theta} = 2.0$, in keeping with the objectives of the data analysis in the next section.

Table 3.4: Table showing variation in h, Type-I error and power across facility size for the continuous and the discrete (OR) CUSUM.

Facility size	Continuous CUSUM			Discrete (OR) CUSUM		
(in transplants/yr)	h	Type-I error	Power	h	Type-I error	Power
10	2.25	0.153	0.761	1.80	0.151	0.664
25	3.00	0.155	0.922	2.50	0.148	0.898
40	3.45	0.151	0.963	2.96	0.151	0.954
60	3.81	0.149	0.990	3.36	0.153	0.984
80	4.20	0.154	0.998	3.65	0.152	0.997
100	4.35	0.154	1.000	3.87	0.155	0.998

From Table 3.4, we see that control limits are a bit higher for the continuous CUSUM and that it yields better power for a similar Type-I error than the discrete (OR) CUSUM. The control limits increase with facility size. This is intuitive to expect as more transplants in 5 years would mean higher arrival rates on average (meaning lower ARLs for the same h, see 3.10, 3.13), thus requiring a higher control limit to achieve similar signal times as smaller facilities. These results are based on a fixed decision period of $T = 5$ years. If the decision boundary T is lowered, then a smaller value of h may suffice and if it is raised, then a larger h may be needed to achieve similar bounds on the Type-I error. Similarly, if the alternative is shifted

towards the right, a larger value of h would be required to achieve the same Type-I error and would also result in higher power.

Sometimes, CUSUM charts may be used to monitor performances at multiple centers which involves multiple testing of the same hypothesis. Consider M different centers and assume that $0 < m_0 < M$ centers are in-control and the rest $M - m_0 = m_1$ are not. The CUSUM is implemented at each center and it is useful to get an idea of the overall rate of false alarms in this procedure. This leads on to concepts such as False Discovery Rates (FDR), an area discussed in the literature by Benjamini and Hochberg (1995) and Storey (2002) among many others. The FDR is defined as the proportion of false discoveries or alarms amongst all discoveries or alarms. The idea of using an FDR approach in the context of multiple CUSUM charts has been proposed by Aylin et al. (2003), Marshall et al. (2004) and Grigg and Spiegelhalter (2005) among many others. Following Marshall et al. (2004) and Storey (2002), we define FDR_T^h, the rate of false discovery by the decision boundary T, as,

$$FDR_T^h = \frac{q_0[\text{Type-I error}]}{q_0[\text{Type-I error}] + (1 - q_0)[\text{Power}]}$$

where $q_0 = m_0/M$ is the proportion of centers that are in-control. We conducted simulation studies where we considered $M = 100$ centers with $m_0 \sim \text{Bin}(M, p_0)$, $T = 5$, different transplant rates $\psi = 10, 40, 60$ and 100, various choices for p_0 as $0.7, 0.8$ and 0.9 and computed Monte-Carlo estimates of FDR_T^h based on the control limits selected in Table 3.4. Once again, we looked at the two CUSUM charts of Sections 3.3 and 3.4.1, respectively. Table 3.5 displays the results, where we see that the FDR increases with the proportion of in-control centers. For the discrete (OR) CUSUM, the FDR is relatively stable across facility size as compared to the continuous CUSUM.

The above approach has the restrictive feature of having to consider a fixed deci-

sion boundary (see, for example, Woodall et al. (2006)), which in effect is similar to a fixed sample size problem and so Grigg and Spiegelhalter (2006) have proposed a different approach for controlling the FDR based on ideas in Benjamini and Hochberg (1995) and studying the stationary distribution of the in-control CUSUM statistic by running it without a control limit. Under the assumption of normally distributed outcomes, the approximate stationary distribution they obtain consists of a mass at zero and a continuous part for values strictly larger than zero.

Table 3.5: Table showing variation in FDR_T^h across h for the continuous and discrete (OR) CUSUM.

Facility size (transplants per year)	Continuous CUSUM				Discrete (OR) CUSUM			
		FDR_T^h				FDR_T^h		
	h	$p_0 = 0.7$	$p_0 = 0.8$	$p_0 = 0.9$	h	$p_0 = 0.7$	$p_0 = 0.8$	$p_0 = 0.9$
10	2.25	0.05	0.44	0.64	1.80	0.35	0.48	0.67
40	3.45	0.12	0.19	0.33	2.96	0.28	0.40	0.60
60	3.81	0.26	0.38	0.58	3.36	0.27	0.38	0.58
100	4.35	0.25	0.37	0.57	3.87	0.27	0.39	0.59

It is helpful to remember here that Table 3.5 and the simulations therein are only for descriptive purposes. From the viewpoint of the SRTR, it is more relevant to control for Type-I error and improve the power of detection (as in Table 3.4) rather than the false discovery rate. The FDR is geared more towards trying to assess the accuracy of the procedure once a center is flagged. Therefore, for purposes of choosing a suitable control limit, we suggest using Table 3.4 as a guideline and Table 3.5 is to be seen merely as a descriptive measure of the procedures discussed.

3.7 Application to kidney transplantation data

The continuous and discrete (OR) CUSUMs were also applied to data from the Scientific Registry of Transplant Recipients (SRTR). The SRTR is administered by Arbor Research Collaborative for Health in collaboration with the University of Michigan. The data includes information on all 59,650 kidney transplants performed

at 258 transplant centers in the United States from January 1997 to December 2001. The primary outcome of interest was graft failure, including death with a functioning graft at 1 year post-transplant. Overall, 5502 (9.22%) transplants failed within 1 year from transplant. The unadjusted failure rate ranged from 0 to 66.7% across the 258 centers.

3.7.1 Risk-adjustment

For the discrete (OR) CUSUM, a logistic regression model was used, adjusting for a number of risk factors (see Table 3.6), to arrive at a risk-adjusted predicted probability of failure $\pi_{i0},\ i \geq 1$ at 1 year post-transplant. This was then used to construct the CUSUM chart for each center separately as indicated in Section 3.4.1 before. We used $OR_A = 2.0$ in order to tune the CUSUM to detect a doubling of the odds of an outcome.

For the continuous CUSUM, a Cox proportional hazards model was used, adjusting for the same risk factors (see Table 3.7), to arrive at an adjusted hazard rate of failure at 1 year post-transplant. This was then used to compute $\alpha_i(x)$ for $i \geq 1$ and incorporated into the CUSUM diagram for each facility as discussed before in Section 3.3. Here also, we worked with $e^\theta = 2.0$ to detect a doubling of the hazard rate of an outcome.

The CUSUM charts were then graphed against time. Figure 3.2 shows the CUSUM curve for a particular facility, where we have added 1 to the timescale to acknowledge the fact that we register the outcome in the CUSUM only after a year has past from the date of transplant. Similar plots are given in Axelrod et al. (2006), but they use transplant number on the horizontal axis. We chose the actual time from transplant as the time scale for comparing with the continuous CUSUM. Figure 3.3 shows the continuous time CUSUM for the same facility, plotted against

Table 3.6: Table of risk factors adjusted for in the logistic model.

Risk Factor	Odds Ratio	P	Risk Factor	Odds Ratio	P
Donor Cause of Death			Congenital/familial disorders	1.04	0.767
Anoxia	0.97	0.599	Diabetes	1.11	0.034
Cerebrovascular/Stroke	1.12	0.011	Renovascular diseases	1.12	0.127
Head Trauma	1.00	Ref	Neoplasms	1.14	0.633
CNS Tumor	0.63	0.027	Hypertensive nephrosclerosis	1.19	0.001
Other	1.10	0.416	Retransplant/graft failure	1.20	0.027
Missing	0.73	0.028	Glomerular disease	1.00	Ref
Donor Age (years)			Other	1.10	0.131
<18	1.08	0.225	Missing	1.17	0.459
18-34	0.89	0.005	Number of B mismatches		
35-49	1.00	Ref	Zero	1.00	Ref
50-64	1.13	0.012	1	1.30	<0.0001
≥ 65	1.43	<0.0001	2	1.40	<0.0001
Donor Race			Number of DR mismatches		
White	1.00	Ref	Zero	1.00	Ref
African-American	1.20	0.0001	1	1.15	0.001
Asian	0.95	0.631	2	1.25	<0.0001
Other	1.03	0.842	Peak PRA		
Donor to Recipient weight ratio			0-9%	1.00	Ref
Q1 (0-0.75)	1.18	0.004	10-79%	1.11	0.007
Q2 (0.75-0.90)	1.20	0.001	>80%	1.34	<0.0001
Q3 (0.90-1.50)	0.99	0.961	PRA Missing	0.93	0.741
Q4 (>1.5)	1.00	Ref	Previous Transplant	1.15	0.053
Missing	1.20	0.001	Recipient BMI		
Recipient Age (years)			<20	1.16	0.035
18-34	1.07	0.144	20-24.9	1.00	Ref
35-49	1.00	Ref	25-29.9	1.06	0.203
50-64	1.38	<0.0001	>30	1.12	0.030
≥ 65	1.43	<0.0001	BMI missing	1.15	0.009
Recipient Ethnicity			Symptomatic PVD	1.22	0.009
Hispanic	0.73	<0.0001	Symptomatic PVD missing	0.95	0.483
Non-Hispanic	1.00	Ref	Dialysis Status		
Missing	1.09	0.446	No dialysis	0.94	0.333
Recipient Race			Peritoneal dialysis	1.09	0.043
White	1.00	Ref	Hemodialysis	1.00	Ref
African-American	0.96	0.282	Unknown type	0.64	0.001
Asian	0.70	<0.0001	Angina/coronary artery disease	1.10	0.045
Other	0.86	0.178	Any previous transfusions	1.12	0.0008
Deceased Donor hypertension	1.12	0.016	Unknown or missing transfusions	1.05	0.248
Expanded criteria donor	1.19	0.007	No previous transfusions	1.00	Ref
Cause of ESRD			Drug treated systemic hypertension	0.90	0.011
Tubular/Interstitial disease	1.18	0.020	Missing	1.01	0.890
Polycystic kidneys	0.80	0.001			

chronological time. Figure 3.5 shows a facility where an abrupt change in the failure rate may have occurred, whereas Figure 3.6 shows one which may be operating at or under the national average rate.

For both of the above risk-adjusted models, we used data pooled from all facilities to get an idea of the national average failure rate. This is then compared to each center via a continuous sequential monitoring scheme. Also, in both the risk-adjustment models, factors like recipient age, donor to recipient weight ratio, donor age and presence of other diseases, among several other factors, are highly predictive of transplant failure.

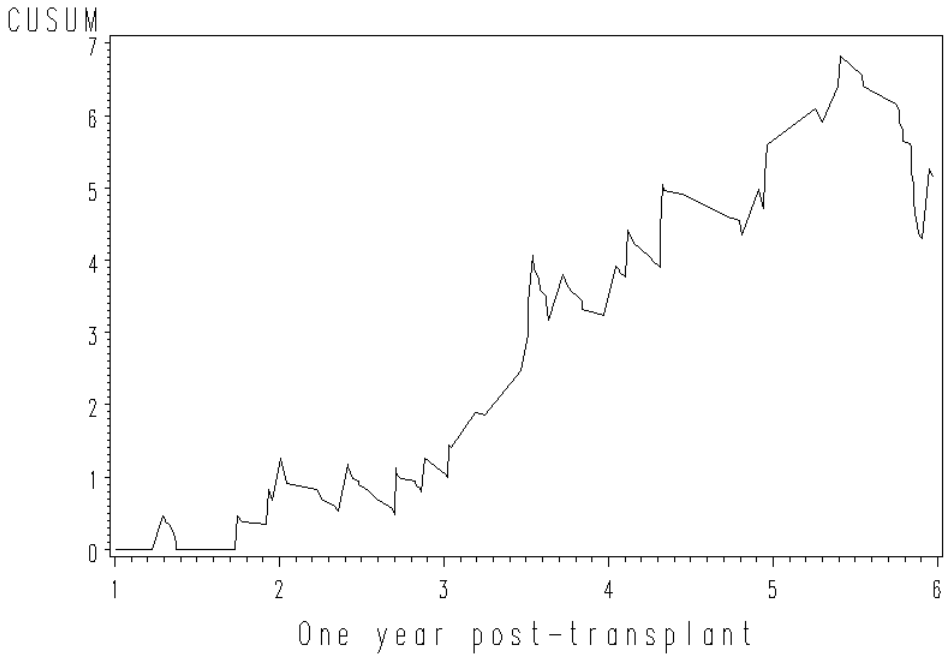

Figure 3.2: The discrete (OR) CUSUM diagram showing a facility possibly operating at a higher failure rate than the national average.

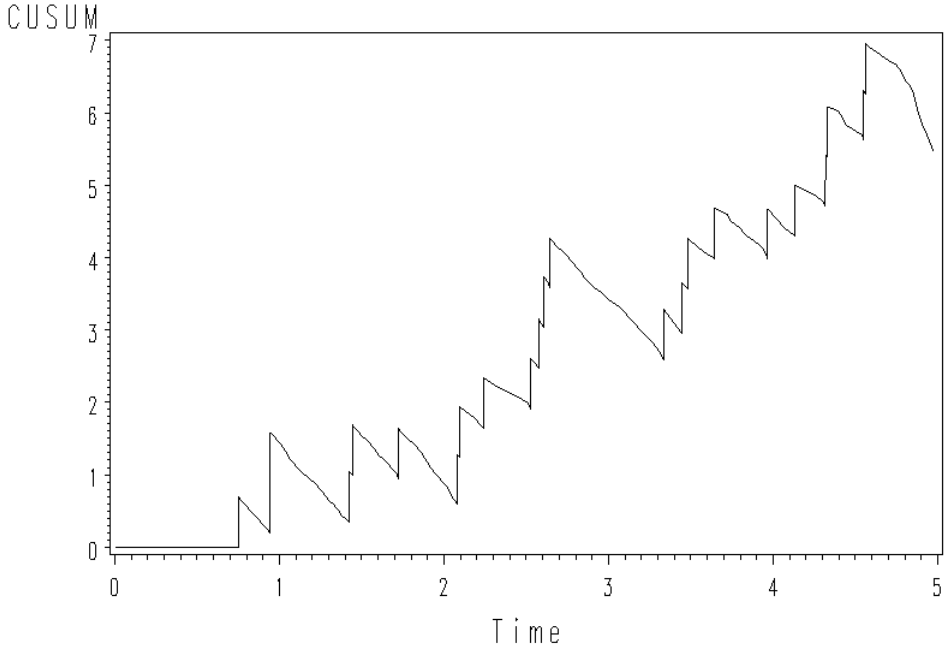

Figure 3.3: The continuous CUSUM diagram showing a facility operating at a possibly higher failure rate than the national average.

Table 3.7: Table of risk factors adjusted for in the Cox model.

Risk Factor	Hazard Ratio	P	Risk Factor	Hazard Ratio	P
Donor Cause of Death			Congenital/familial disorders	1.17	0.055
Anoxia	1.02	0.560	Diabetes	1.14	<0.0001
Cerebrovascular/Stroke	1.09	0.002	Renovascular diseases	1.14	0.004
Head Trauma	1.00	Ref	Neoplasms	1.14	0.434
CNS Tumor	0.88	0.284	Hypertensive nephrosclerosis	1.18	<0.0001
Other	0.94	0.436	Retransplant/graft failure	0.94	0.270
Missing	0.63	<0.0001	Glomerular disease	1.00	Ref
Donor Age (years)			Other	1.01	0.889
<18	0.95	0.187	Missing	0.98	0.888
18-34	0.90	<0.0001	Number of B mismatches		
35-49	1.00	Ref	Zero	1.00	Ref
50-64	1.16	<0.0001	1	1.13	0.001
$\geq$ 65	1.39	<0.0001	2	1.16	<0.0001
Donor Race			Number of DR mismatches		
White	1.00	Ref	Zero	1.00	Ref
African-American	1.15	<0.0001	1	1.13	<0.0001
Asian	0.96	0.516	2	1.18	<0.0001
Other	1.06	0.546	Peak PRA		
Donor to Recipient weight ratio			0-9%	1.00	Ref
Q1 (0-0.75)	1.19	<0.0001	10-79%	1.10	<0.0001
Q2 (0.75-0.90)	1.14	0.0004	>80%	1.29	<0.0001
Q3 (0.90-1.50)	1.02	0.497	PRA Missing	1.07	0.554
Q4 (>1.5)	1.00	Ref	Previous Transplant	1.33	<0.0001
Missing	1.58	<0.0001	Recipient BMI		
Recipient Age (years)			<20	1.22	<0.0001
18-34	1.17	<0.0001	20-24.9	1.00	Ref
35-49	1.00	Ref	25-29.9	0.99	0.837
50-64	1.15	<0.0001	>30	1.02	0.650
$\geq$ 65	1.47	<0.0001	BMI missing	1.04	0.247
Recipient Ethnicity			Symptomatic PVD	1.23	<0.0001
Hispanic	0.79	<0.0001	Symptomatic PVD missing	0.92	0.106
Non-Hispanic	1.00	Ref	Dialysis Status		
Missing	1.19	0.010	No dialysis	0.85	0.0001
Recipient Race			Peritoneal dialysis	1.02	0.353
White	1.00	Ref	Hemodialysis	1.00	Ref
African-American	1.11	<0.0001	Unknown type	0.55	<0.0001
Asian	0.72	<0.0001	Angina/coronary artery disease	1.16	<0.0001
Other	0.81	0.005	Any previous transfusions	1.18	<0.0001
Deceased Donor hypertension	1.08	0.008	Unknown or missing transfusions	1.08	0.002
Expanded criteria donor	1.15	0.0005	No previous transfusions	1.00	Ref
Cause of ESRD			Drug treated systemic hypertension	0.97	0.349
Tubular/Interstitial disease	1.09	0.049	Missing	1.08	0.169
Polycystic kidneys	0.77	<0.0001			

3.7.2 Choosing a control limit

In this section, we note some issues in deciding on a control limit for the CUSUM charts constructed for the above data. Figure 3.4 shows the proportion signaled by the two CUSUM charts across different choices of control limit. For this figure, a single control limit has been chosen for all the facilities. We see that the fraction signaled is similar for both the charts, with the discrete CUSUM signaling slightly more often than the continuous CUSUM for lower control limits. However, choosing a single control limit may not be the right thing to do here as can also be seen from Table 3.4.

Note that, here we can also think about a fixed decision boundary problem as in Section 3.6, with $T = 5$. The 258 transplant facilities are of widely different sizes with number of transplants ranging from 1 to 1533 over the 5 year period. Since the number of transplants until a signal and the Type-I error depend on the volume of transplants at a facility, we grouped the 258 facilities into 5 categories based on the total number of transplants performed in 5 years. There were 50 facilities performing less than 8 transplants per year on average and 28 facilities performing more than 100 transplants per year on average. For very small facilities, the CUSUMs may not yield much power, so we left these 50 small facilities out of the analysis. The 28 facilities with more than 100 transplants per year vary widely in terms of number of transplants per year with the maximum being 306 transplants per year and therefore were not included in the analysis either. The remaining 180 facilities were divided in 3 categories and different control limits were fixed for each of these categories using the numbers from Table 3.4.

Table 3.8 shows the control limits for the 3 categories along with the number of facilities flagged in each category based on the continuous and the discrete (OR)

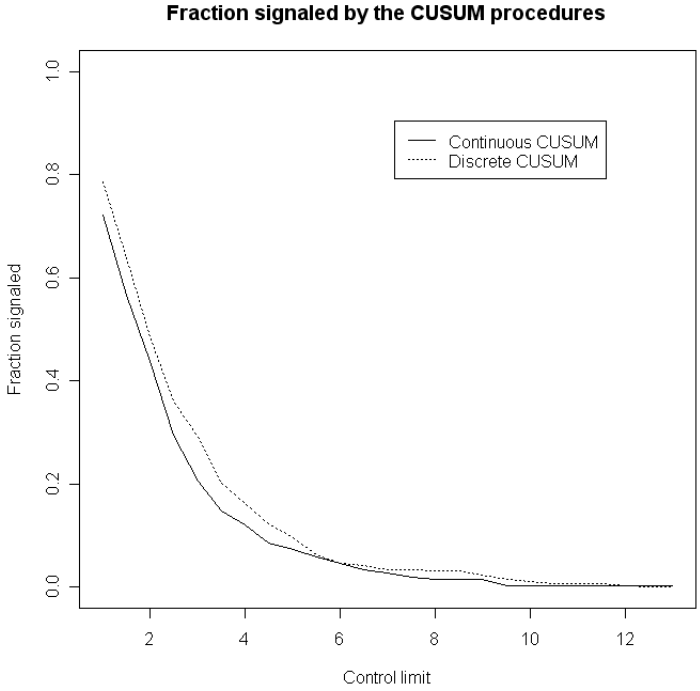

Figure 3.4: Fraction of facilities signaled by the two CUSUM charts when a single control limit is used.

CUSUM. In all, 37 facilities were flagged by the continuous CUSUM and 41 facilities were flagged by the discrete (OR) CUSUM. For this dissertation, the analysis was done with facility identity blinded and so we do not identify which facilities are flagged.

Table 3.8: Table showing number of facilities flagged by the continuous and discrete (OR) CUSUMs.

Facility size (in transplants/yr)	Total number of facilities	Continuous CUSUM		Discrete (OR) CUSUM	
		Control limit h	Number flagged	Control limit h	Number flagged
8 - 20	33	2.25	7	1.80	8
20 - 50	89	3.45	16	2.96	19
50 - 100	58	4.20	14	3.65	14

In summary, the choice of a control limit for a certain facility should depend on the length of follow-up intended for that institution, the average rate of transplants at that facility and the level of Type-I error to be controlled for. Based on this information, a control limit may be chosen based on simulating the CUSUM under

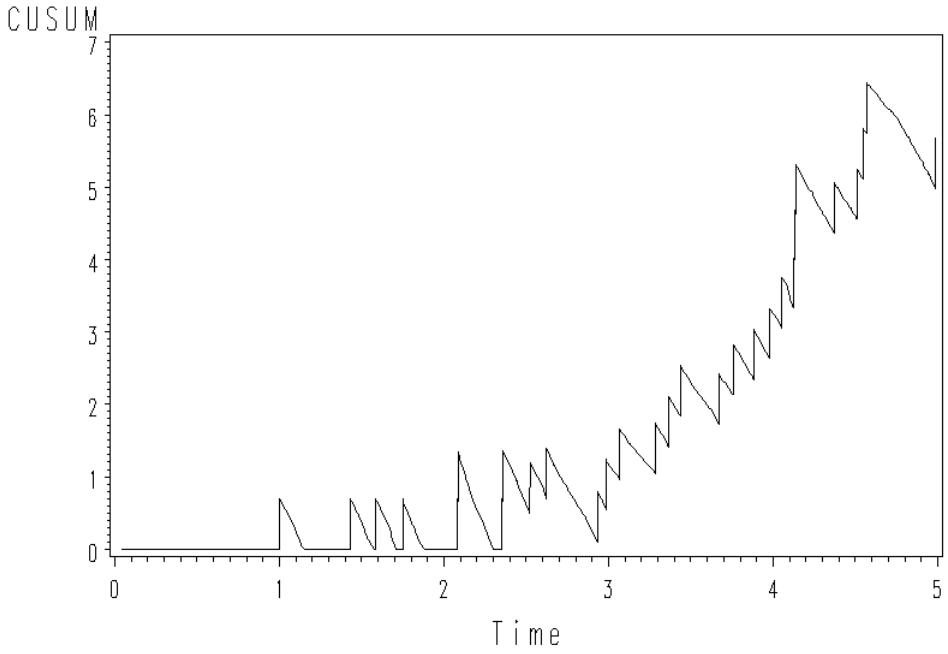

Figure 3.5: The continuous CUSUM diagram showing a sudden possible change in the failure rate.

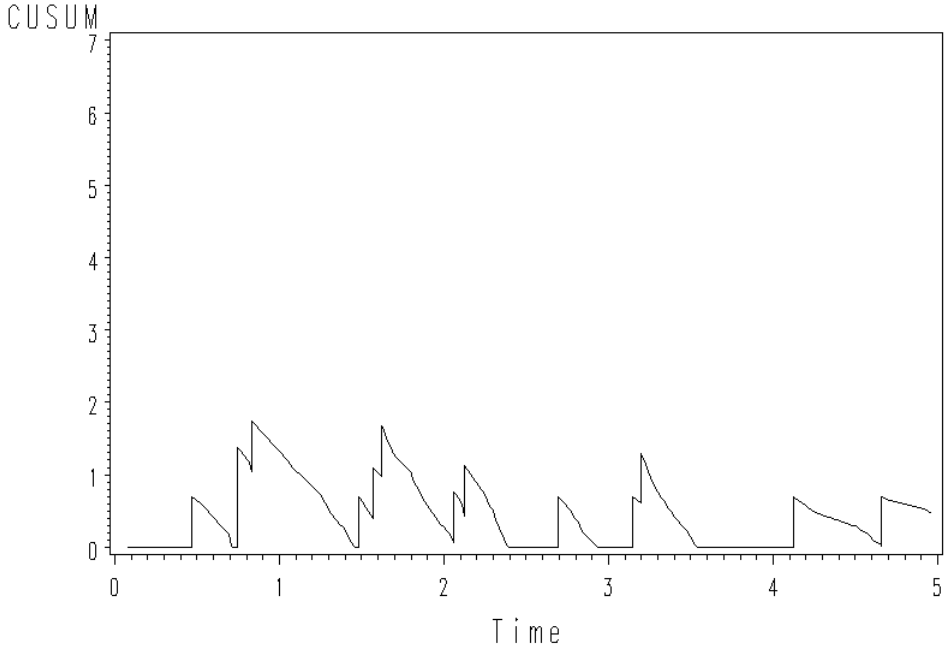

Figure 3.6: The continuous CUSUM diagram showing a facility operating near the national average failure rate.

the null as in Section 3.6 and using a calibration strategy as in Table 3.4. For small centers, having less than 8 or 10 transplants per year on average, it is difficult to choose a control limit that keeps the Type-I error low as well as yield good power under alternatives not so far away from the null. The analogy with a fixed sample size test becomes clearer when the follow-up is fixed and it is common knowledge that one needs higher sample sizes to detect alternatives close to the null with greater certainty. Therefore, for smaller centers, increasing the follow-up or time window might be a possible answer.

3.8 Discussion

A continuous time version of the CUSUM is introduced which, compared to earlier proposals, utilizes information available from the data in a more timely manner. It is simple to compute and results in substantial savings in the average run lengths to provide a signal to institutions with higher failure rates. This process has been compared both analytically and numerically with discrete time analogues.

Theoretical approximations to the mean run length have been derived using Wald's identity. While the approximation works well when the drift parameter $\eta > 0$, it does not work so well for $\eta < 0$. To assess run lengths for smaller values of η, simulation is probably the best approach. From a practical point of view, however, early detection and accuracy is important when $\eta > 0$. When $\eta < 0$, both the empirical estimate and theoretical approximation of the mean run length are large and errors in approximations are less important. Adjustments for ignoring the overshoot in applying Wald's identity which are discussed in Khan (1978) are complex and as we see from Table 3.2, the error in approximation diminishes with increasing h. Exact expressions for the mean run length using the Poisson likelihood ratio incre-

ment in (3.2) have been derived by Dvoretzky et al. (1953) and Zacks (2004) but the formulas therein are very involved and computationally intensive.

Other approaches have also been suggested in the literature for deriving approximations to the ARL. Methods using integral equations have been discussed in Goel and Wu (1971). Approximations using Markov chain methods have been discussed by Brook and Evans (1972) and Steiner et al. (2000). Ewan and Kemp (1960) and Woodall (1983) discuss an approach using recursive equations and Reynolds (1975) discusses a Brownian motion approximation. Approximations in the case of Exponentially distributed outcomes are discussed in Gan (1992) and Vardeman and Ray (1985). Some approximations using Wald's identity have been discussed in Siegmund (1985), where a different expression is obtained and the author mentions that they do not work well. It is of some interest to investigate if improvements on the current approximations are possible for the case of a negative drift ($\eta < 0$) using some of the above methods. The simulations and the approximations in Sections 3.3.3 and 3.5, are obtained by starting the CUSUM at a year after $t = 0$ when the transplant process $N^A(t)$ is in equilibrium. In principle, one can also start the procedure at $t = 0$ and empirical evidence shows that the continuous CUSUM is still doing better than the discrete versions. However, theoretical approximations to the ARL would be much harder to obtain in that case.

The distribution of the covariates at the facility of interest may alter the ARL to some extent if the 1 year failure probability at the facility changes substantially compared to that in the population. However, as Table 3.3 suggests, such alterations are largely due to location effects in the covariate distribution and one may use the moment-generating function of the covariate distribution to compute approximate factors by which the theoretical ARLs will change.

In practice, the choice of a control limit for a particular facility in multi-center studies demands further research. For small centers, it is difficult to choose a control limit that controls the Type-I error as well as yield good power. In this chapter, we have allowed the Type-I error to be 15% which may not be unreasonable for small centers, given that it would be less costly to review them compared to bigger centers. So, depending on the goals of the study, different control limits may be chosen.

In the implementation of CUSUM procedures, it is important to remember that CUSUM signaling does not necessarily prove a clinically important decline or improvement in clinical quality has occurred. Rather, the signal suggests that closer examination by the quality improvement team may be required. Although the CUSUM has been set up as a test of hypothesis problem, this is an important philosophical distinction from the usual accept or reject outcome of a test.

The method of likelihood ratio scoring may be applied to several settings that involve monitoring outcomes. In this chapter we have used a binary outcome and a Poisson approximation. However, one can consider different outcomes and a corresponding model to form the likelihood score at each time point and define a CUSUM process in a similar fashion. Theoretical approximations to the average time to signal may also be obtained similarly by exploiting the renewal nature of the CUSUM process and employing Wald's identity.

The continuous time CUSUM procedure may be used in any situation that requires constant surveillance and monitoring. It is important to note, however, that the CUSUM then also needs to incorporate information about outcomes as soon as they occur. Sometimes, due to organizational or design issues, there can be a reporting delay between the actual occurrence of the event and the time it gets reported. The delay may be random or deterministic, depending on the study design. It is

of considerable interest and importance to incorporate this delay feature into the CUSUM procedure.

Acknowledgements

We wish to thank Mary K. Guidinger from Arbor Research Collaborative for Health and Randall L. Webb from The University of Michigan Kidney Epidemiology and Cost Center for helpful discussions and assistance with organizing the computational results from SRTR data. We would also like to thank Dr. Robert M. Merion and Dr. David A. Axelrod for introducing the problem to us and for helpful inputs and discussions.

The Scientific Registry of Transplant Recipients is funded by contract number 234-2005-37009C from the Health Resources and Services Administration, US Department of Health and Human Services. The views expressed herein are those of the authors and not necessarily those of the US Government. This study was approved by HRSA's SRTR project officer. HRSA has determined that this study satisfies the criteria for the IRB exemption described in the 'Public Benefit and Service Program' provisions of 45 CFR 46.101(b)(5) and HRSA Circular 03.

CHAPTER IV

A CUSUM incorporating reporting delays

4.1 Introduction

In multi-center studies involving an intervention that generates an outcome as time progresses, it is useful to be able to monitor these outcomes continuously in real-time and raise an alarm if a significant upward trend is detected in the failure rate of the intervention. The CUSUM technique, introduced in Page (1954), has been used as a graphical sequential monitoring scheme for some time now. It has been used in the context of monitoring organ transplantation outcomes in Axelrod et al. (2006) and Steiner et al. (2001).

Similar to any real-time monitoring scheme, an effective implementation of the CUSUM procedure also demands the availability of outcome information on a real-time basis. Stated simply, outcome information needs to be reported to the CUSUM as soon as it occurs. In actual practice, however, outcome reporting is seldom instantaneous. In this chapter, we devise a CUSUM procedure when there is a delay involved in the reporting of outcomes. We assume that the outcome is reported only after a random delay period has elapsed from the time it occurred.

It is helpful here to consider again the example of kidney transplants. Suppose, there is a facility or institution that performs organ transplants over time. Some of

these transplants might fail, either due to the death of the patient or due to graft failure. Over time, this institution will accumulate transplant failures and in the interest of quality, it is important to raise an alarm if a significant upward trend in the failure rate is detected. Individual failure times will also be affected by covariate information and so we need to look at a risk-adjusted procedure. In this chapter, we shall construct a risk-adjusted CUSUM technique under the assumption that conditional on the time of transplant, the transplant failure time and the delay time are jointly independent.

4.2 A CUSUM that incorporates reporting delays

In this section, we shall define a CUSUM procedure in continuous time for a particular institution that accounts for the reporting delay mechanism. Risk factors that may affect individual outcomes are also adjusted for.

Assume that we have a very large population arising from the combined data of many institutions with information on times of transplant, post-transplant failure times and delays involved in the reporting of the post-transplant failure. As has been considered in Axelrod et al. (2006) before, we shall assess transplant outcomes based on a one-year performance. We define a qualifying failure as one where the transplant fails within a year from the date of transplant with the failure time being measured from the date of transplant. Let $f_0(x|Z),\ x > 0$ be the conditional density (defective) for a qualifying failure time, given a covariate value Z. Note that $f_0(x|Z) = 0$ for all $x > 1$ and that $F_0(1|Z) = F_0(\infty|Z) < 1$ is the probability of a failure in the first year. Assume that we can estimate $f_0(x|Z)$ from a regression model, adjusting for covariates Z, based on all the data available in this population. Let $h_0(l),\ l > 0$ be the density for the reporting delay times, measured from the time the transplant

fails. Assume that we can also estimate $h_0(l)$ from this population via a regression model. For the present application, we shall ignore the uncertainty in the estimates $f_0(x|Z)$ and $h_0(l)$ as seems reasonable since these estimates are based on a large number of observations from the population.

For a certain institution of interest, assume that $S_1, S_2, \ldots$ represent the times of transplant, that is, the i-th transplant takes place at time S_i. Later on we shall assume that $S_1, S_2, \ldots$ are the arrivals of a Poisson process, but for the moment we shall work conditionally on the times.

Let $X_1, X_2, \ldots$ be the qualifying post-transplant failure times with $Z_1, Z_2, \ldots$ being the corresponding covariate values measured after transplant. It is of primary interest to ascertain if the probability of a qualifying failure at this particular institution, conditional on covariate values, is higher or lower than that at the population discussed before. One simple parametric model that relates the failure experiences at this institution to that at the population is the proportional density model as,

$$f_{\mu,i}(x|Z_i) = e^{\mu} f_0(x|Z_i) \tag{4.1}$$

where, $f_{\mu,i}(x|Z_i),\ x > 0$ is the conditional density (defective) of X_i given Z_i. Here, μ can be any real number depending on whether the institution has a better ($\mu < 0$) or worse ($\mu > 0$) failure experience than the population. Observe that this model is similar to a proportional hazards model when the probability of a one-year failure $F_0(1|Z)$ is small.

Let $L_1, L_2, \ldots$ be the delay times with L_i being the reporting delay corresponding to the i-th transplant. We assume that L_i is measured from the time (chronological) of failure, $S_i + X_i$ and has density $h_0(l),\ l > 0$ which is the same as that in the population. Typically, each institution will have a different rate of reporting but, for the moment we shall assume that this institution has a similar delay distribution

as the population. The overall structure of the random variables so defined is that given $S_1, S_2, \ldots,\ (X_i, L_i, Z_i),\ i = 1, 2, \ldots$ are iid and independent of the $\{S_i\}$.

Let $T_i^* = S_i + X_i + L_i$ be the chronological time of report and define $N_i(t) = I(T_i^* \leq t)$ as the process indicating if the outcome of the i-th transplant has been reported or not by time t. It is instructive here, to note the geometry (see Figure 4.1) of the region where we observe the data. Having observed the time S_i of the i-th transplant, we receive its report by time t ($N_i(t) = 1$) if and only if $X_i + L_i \leq t - S_i$. In other words, $N_i(t) = 1$ if and only if (X_i, L_i) lies in the triangular region in bold in Figure 4.1.

We construct the likelihood $L_t(\mu)$ based on the data available upto time t. We shall base our likelihood on the counting process $N_i(t),\ i \geq 1$, the reason for which will become clear presently. We have,

$$L_t(\mu) \propto \prod_{i \geq 1} [e^{\mu} f_0(x_i|z_i) h_0(l_i)]^{N_i(t)} [1 - e^{\mu} B_i(t)]^{1 - N_i(t)}$$

where

$$B_i(t) = \int_0^{t - S_i} f_0(x|z_i) H_0(t - S_i - x)\ dx$$

and we define $B_i(t) = 0$ if $t < S_i$. Note that, $e^{\mu} B_i(t) = E[N_i(t)|S_i]$. From this, it can be easily seen that the likelihood ratio comparing two values of μ depends only on the process $N_i(t),\ i \geq 1$ and not on the observed values of X_i and L_i. This is a feature of the proportional density model (4.1) and is the main motivation for it. Also, a factorization argument shows that $L_t(\mu)$ depends on μ only through $N(t) = \sum_{i \geq 1} N_i(t)$. This may also be interpreted as a sufficiency property of the process $N(t)$ for μ.

We now construct a CUSUM based on a SPRT, introduced by Wald (1947), defined at each time point. For this purpose, we select a constant $\theta > 0$ and consider

a likelihood ratio test of $\mu = \theta$ versus $\mu = 0$ using the likelihood $L_t(\mu)$. The constant θ is chosen in such a way as to design the CUSUM to detect failure rate changes for μ in the range $[0, \theta]$. The case $\mu = 0$ represents the population failure rate and it is of main interest to discover an upward trend in failure rates from the population. Therefore, the log-likelihood difference at time t is given by,

$$\begin{aligned} U_t &= \log L_t(\theta) - \log L_t(0) \\ &= \sum_{i \geq 1} \left[\theta N_i(t) + (1 - N_i(t))\{\log(1 - e^{\theta} B_i(t)) - \log(1 - B_i(t))\}\right] \end{aligned} \tag{4.2}$$

In defining the CUSUM, it is convenient to work with the increments dU_t (of U_t) and we find that,

$$dU_t = dN^*(t) - (e^{\theta} - 1) dA(t) \tag{4.3}$$

where,

$$\begin{aligned} dN^*(t) &= \sum_{i \geq 1} \left(\theta - \log \frac{1 - e^{\theta} B_i(t)}{1 - B_i(t)}\right) dN_i(t) \\ dA(t) &= \sum_{i \geq 1} \frac{(1 - N_i(t)) B_i'(t)}{[1 - e^{\theta} B_i(t)][1 - B_i(t)]}\, dt \end{aligned}$$

Since we are considering a one-sided alternative and are interested in detecting upward trends only in the CUSUM, we define a one-sided CUSUM $\{G_t\}$, with $G_0 = 0$ as,

$$G_{t+dt} = \max(0, G_t + dU_t)$$

or equivalently as,

$$G_t = U_t - \min_{0 \leq s \leq t} U_s$$

In order to implement the CUSUM procedure, we track the process G_t continuously in time and say that the CUSUM has raised an alarm at time τ_h, where,

$$\tau_h = \inf\{t > 0 : G_t \geq h\},$$

the first time when G_t crosses a fixed upper barrier $h > 0$.

4.3 The case of Poisson arrivals

In this section, we shall investigate theoretical properties of $E(\tau_h)$ under the assumption that the arrivals or transplants at the facility occur in accordance to a homogeneous Poisson process. For the CUSUM defined in (4.3), the increments are complex quantities and it is difficult to obtain a theoretical approximation to the ARL using Wald's identity applied directly for these increments. Here, we shall obtain a slightly simpler likelihood under the homogeneous Poisson process assumption. We shall also assume that the covariates at the institution are homogeneous and so we can work with the marginal distribution of the post-transplant failure times.

Consider first, the case when time is discrete and measured on the integers as $t = 1, 2, \ldots$. Let $R_s \geq 0$ be the number of transplants at a facility at time s. Assume that R_s is Poisson distributed with mean ψ. Also, as before, we have a qualifying post-transplant failure time X having a (defective) distribution as,

$$f(x) = e^{\mu} f_0(x)$$

for $x = 0, 1, \ldots, a$ and zero otherwise. Here, a is a time period for defining a qualifying failure and usually $a = 365$. We also have a random delay time L taking values $0, 1, 2, \ldots$ and having a density $h_0(l)$, $l = 0, 1, 2, \ldots$. We also define N_{sxl} as the number of transplants that have occurred at time s, qualifiably failed at time $s + x$ and reported at time $s + x + l$. Note that we can observe N_{sxl} by time t only if $x + l \leq t - s$. Also, let

$$\begin{aligned} N_s^t &= \sum_{x+l \leq t-s} N_{sxl} \qquad \text{and,} \\ \rho_s^t &= \sum_{x+l \leq t-s} \psi f_0(x) h_0(l) \end{aligned}$$

The data available up to time t is $\{N_{sxl} : x + l \leq t - s, t - s \geq 0\}$ and $\{R_s - N_s^t :$

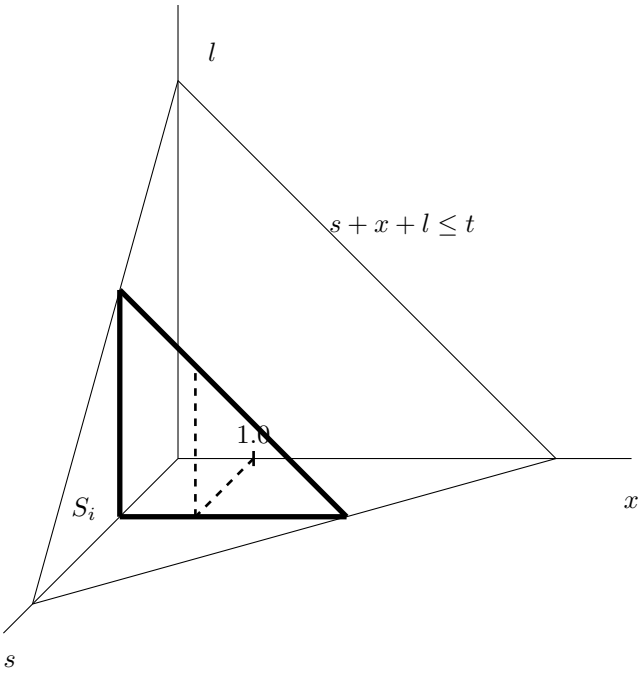

Figure 4.1: Geometry of the delay problem.

$t - s \geq 0\}$. The likelihood of μ based on the data available up to time t is,

$$L_t(\mu) \propto \prod_{s=1}^{t} \left[\prod_{x+l \leq t-s} [e^{\mu} f_0(x) h_0(l)]^{N_{sxl}} [1 - e^{\mu} \rho_s^t]^{R_s - N_s^t} \right]$$

Observe that this likelihood implicitly conditions on the time of transplant being equal to s. We can compute the expected Fisher information $I(\mu)$ for μ from this likelihood as,

$$I(\mu) = \sum_{s=1}^{t} \frac{e^{\mu} \rho_s^t}{1 - e^{\mu} \rho_s^t}$$

Observe now that N_{sxl} may be interpreted as a Poisson process in three dimensions. For every time t, define the set $Q_t = \{(s, x, l) : s + x + l \leq t\}$ and let

$$N_t = \sum_{(s,x,l) \in Q_t} N_{sxl}$$

Under the assumption that the number of transplants $R_s,\ s \geq 1$ are Poisson distributed with rate ψ, one can show that N_t is also Poisson distributed with rate $e^{\mu} \rho_t$,

where

$$\begin{aligned}
\rho_t &= \sum_{(s,x,l)\in Q_t} \psi f_0(x) h_0(l) \\
&= \sum_{s=1}^{t} \sum_{x=0}^{t-s} \sum_{l=0}^{t-s-x} \psi f_0(x) h_0(l) \\
&= \sum_{s=1}^{t} \sum_{x=0}^{t-s} \psi f_0(x) H_0(t-s-x) \\
&= \psi \sum_{x=0}^{t-1} f_0(x) \sum_{s=1}^{t-x} H_0(t-x-s) \\
&= \psi \sum_{x=0}^{t-1} f_0(x) \sum_{l=0}^{t-1-x} H_0(l) \qquad (4.4)
\end{aligned}$$

Therefore, the likelihood of μ based on N_t is,

$$L_t^*(\mu) \propto [e^{\mu}\rho_t]^{N_t} e^{-e^{\mu}\rho_t}$$

whereby, the expected Fisher information $I^*(\mu)$ for μ is,

$$I^*(\mu) = e^{\mu}\rho_t$$

If we can show that $I(\mu)$ and $I^*(\mu)$ do not differ much, then we can use $L_t^*(\mu)$ for constructing a CUSUM which will have a much simpler structure than that based on $L_t(\mu)$ in (4.3). We can then derive a theoretical approximation to the ARL for this CUSUM and compare this, using simulations, to the actual implementation as

defined in (4.3). We have,

$$\begin{aligned} I(\mu) &= \sum_{s=1}^{t} \frac{e^{\mu}\rho_s^t}{1-e^{\mu}\rho_s^t} \\ &= \sum_{s=1}^{t} \left(\frac{1}{1-e^{\mu}\rho_s^t} - 1 \right) \\ &\approx \sum_{s=1}^{t} \left(1 + e^{\mu}\rho_s^t - 1 \right) \\ &= e^{\mu} \sum_{s=1}^{t} \rho_s^t \\ &= e^{\mu}\psi \sum_{s=1}^{t} \sum_{x=0}^{t-s} f_0(x) H_0(t-s-x) \\ &= I^*(\mu) \end{aligned}$$

Here, we have made the approximation $(1-e^{\mu}\rho_s^t)^{-1} \approx 1 + e^{\mu}\rho_s^t$ in view of the fact that $\rho_s^t \leq F_0(t-s) \leq F_0(a)$. The quantity $F_0(a)$ refers to the probability of a failure within a days from transplant in the entire population across all the institutions. In applications, this is usually quite small, around 10% and so this approximation is a reasonable one.

Therefore, we can use $L_t^*(\mu)$ as the likelihood for constructing the CUSUM. As before, we consider the likelihood ratio of $\mu = \theta$ versus $\mu = 0$ using $L_t^*(\mu)$ and obtain,

$$V_t = \theta N_t - (e^{\theta} - 1)\rho_t$$

It is helpful to consider increments in the interval $(t-1, t]$ in the log-likelihood for defining the CUSUM and we obtain

$$\Delta V_t = \theta \Delta N_t - (e^{\theta} - 1)\Delta\rho_t$$

as the increment in the log-likelihood at time t. We shall obtain a theoretical approximation to the ARL of a CUSUM corresponding to the increments at time t being ΔV_t and compare this approximation to the actual ARL obtained from the

CUSUM based on (4.3) in simulations. One can easily show (from (4.4)) that if the delay variable L takes values in a finite range $0, 1, \ldots, K$, say, then $\Delta\rho_t$ is a constant for $t > K + a$ and then we may say that the process N_t is in equilibrium. Since $f_0(x) = 0$ for $x > a$ and $H_0(l) = 1$ for $l > K$, so for $t > K + a$, $\Delta\rho_t = \psi F_0(a)$. Let $\gamma^* = \psi F_0(a)$. Therefore, for $t > K + a$, the increments ΔV_t in the CUSUM become iid and V_t may be treated as a random walk.

We, therefore, consider the CUSUM for $t > K + a$ and work with the definition,

$$V_t = \theta N_t - (e^\theta - 1)\gamma^* t$$

where N_t is Poisson distributed with mean $e^\mu \gamma^* t$.

The above arguments extend directly to the case when time is continuous and is measured in years. Then $V_t = \theta N(t) - (e^\theta - 1)\gamma^* t$ with $N(t)$ being a homogeneous Poisson process with rate $e^\mu \gamma^*$, where $\gamma^* = \psi F_0(1)$, when we consider monitoring V_t for $t > K + 1$. When we consider $N(t)$ in equilibrium, it is possible to obtain approximate expressions for the ARL using Wald's identity for the CUSUM based on V_t. Following identical arguments as in Section 3.3.3, we obtain, therefore,

$$E(\tau_h) \approx \begin{cases} \frac{h}{\eta^*} - \frac{e^{-\mu}(e^\theta - 1)}{\eta^*}\left(\frac{1 - e^{-\omega_0 h}}{1 - e^{-\omega_0 \theta}}\right), & \eta^* \neq 0 \\ \\ \frac{h^2 e^{-\mu}}{\theta^2 \gamma^*}, & \eta^* = 0. \end{cases} \tag{4.5}$$

In the above formulas, η^* represents the drift parameter and is easily seen to be,

$$\eta^* = (\theta e^\mu - e^\theta + 1)\gamma^*.$$

Also, the quantity ω_0 is the non-zero solution to,

$$\exp\{\gamma^* t[\omega(e^\theta - 1) + e^\mu(e^{-\omega\theta} - 1)]\} = 1.$$

Observe that, although the approximations look identical to (3.10), since $\gamma^* \neq \gamma$ in general, the numbers from the formulas in (4.5) will generally be different from those based on the formulas in (3.10). The difference arises from the fact that in the expression for γ obtained in Section 3.3.2, $F_\mu(1|Z_i)$ is derived based on a proportional hazards assumption, whereas in the present situation, $F_\mu(1) = e^\mu F_0(1)$, by (4.1).

However, under equilibrium, the distribution of $N(t)$, and hence that of V_t is free of the delay distribution. The effect of a delay would be observed only if we start monitoring from $t = 0$ and also take into consideration the part where $N(t)$ is not in equilibrium. Theoretical approximations to the ARL would be harder to obtain in that case.

4.4 Simulation study

A simulation study was conducted where transplants occur according to a homogeneous Poisson process with rate $\psi = 100$ transplants per year. The failure time distribution at the population $f_0(x|Z)$ was taken as Exponential (defective) as,

$$f_0(x|Z) = \begin{cases} \lambda_0 e^{-\lambda_0 x}, & 0 < x < 1 \\ 0, & x > 1 \end{cases}$$

where λ_0 is chosen to satisfy

$$1 - e^{-\lambda_0} = 10\%$$

The delay distribution was taken as a two point distribution as,

$$h_0(l) = (1 - \alpha)I(l = 0) + \alpha I(l = 5)$$

where $0 \leq \alpha \leq 1$ is a pre-specified constant controlling the fraction of delayed reports. The CUSUM G_t was simulated based on the random walk U_t in (4.2) wherein we also

Table 4.1: Table showing ARLs for the delay CUSUM simulated under equilibrium conditions.

e^{μ}	Ave. ARL	Med. ARL	SD(ARL)	$E(\tau_h)$
		$\alpha = 0.0$		
1.00	25.91	17.76	25.65	23.59
1.22	6.82	4.70	6.50	6.53
1.50	2.14	1.62	1.72	2.37
1.73	1.22	0.97	0.88	1.39
2.00	0.83	0.69	0.55	0.92
		$\alpha = 0.5$		
1.00	26.08	18.53	26.37	23.59
1.22	6.90	4.92	6.87	6.53
1.50	2.25	1.69	1.88	2.37
1.73	1.36	1.10	1.00	1.39
2.00	0.81	0.70	0.53	0.92
		$\alpha = 1.0$		
1.00	24.83	17.13	24.98	23.59
1.22	7.07	5.18	6.58	6.53
1.50	2.31	1.80	1.88	2.37
1.73	1.26	1.01	0.88	1.39
2.00	0.79	0.67	0.48	0.92

used $e^{\theta} = 2.0$ to detect a shift of 2.0 times the null hazard rate. The control limit was taken as $h = 4.35$ as in Table 3.1 in Chapter III. In Table 4.1, we investigate the cases $\alpha = 0, 0.5$ and 1. We find that the ARLs are quite similar across the three cases. This is because we monitored the CUSUM process beyond $t = 6$ years, when the process $N(t)$ is in equilibrium and then, the rate of the process $N(t)$ is free of the delay distribution. The last column of Table 4.1 also shows the theoretical approximations (4.5) to the ARL and it is clear that they agree quite well with the simulated numbers across all ranges of e^{μ}. The ARLs for the case of no delay are a bit shorter than those obtained in Chapter III for the same control limit and this may be a feature of the proportional density model used here.

Table 4.2: Table showing ARLs for the delay CUSUM simulated from startup ($t = 0$).

	Constant delays			Exponential(1) delays		
e^{μ}	Ave. ARL	Med. ARL	SD(ARL)	Ave. ARL	Med. ARL	SD(ARL)
			$\alpha = 0.0$			
1.00	24.47	18.61	22.21	24.16	18.02	20.91
1.22	7.14	5.27	6.24	6.86	4.98	6.05
1.50	2.70	2.15	1.81	2.69	2.13	1.85
1.73	1.83	1.53	1.00	1.78	1.57	0.93
2.00	1.33	1.22	0.53	1.35	1.26	0.52
			$\alpha = 0.5$			
1.00	26.81	20.07	21.33	22.97	16.99	19.67
1.22	9.34	7.77	6.34	7.55	5.76	6.07
1.50	4.44	3.89	2.51	3.33	2.81	2.09
1.73	3.16	2.75	1.61	2.15	1.93	0.97
2.00	2.20	1.92	1.09	1.66	1.52	0.63
			$\alpha = 1.0$			
1.00	28.13	21.82	20.28	24.91	18.76	20.72
1.22	12.09	9.88	6.68	8.73	6.81	6.99
1.50	7.74	7.18	1.86	3.70	3.19	2.00
1.73	6.75	6.56	0.84	2.68	2.45	1.09
2.00	6.33	6.22	0.52	2.10	1.99	0.64

We also conducted simulations using the same delay distribution as above but

monitoring the CUSUM from the startup period of $t = 0$. While theoretical approximations are not available for this approach, Table 4.2 nevertheless shows that considering the process outside equilibrium results in longer ARLs for all the three cases $\alpha = 0, 0.5$ and 1. An interesting observation is that the ARLs for the case $\alpha = 1$, when there is a constant delay of 5 years for all subjects, are just a translation by exactly 5 years of the ARLs under $\alpha = 0$ (with no delay). The fact that the run length distribution just shifts by a constant is apparent from the fact that the standard deviations remain the same for $\alpha = 0, 1$. It is harder to interpret the numbers for the case $\alpha = 0.5$ as the delays then are a mixture of 0 and 5 years. Table 4.2 also displays similar statistics for Exponential(1) delay times. Observe that since Exponential(1) random variables are not finitely supported, the process $N(t)$ is never in equilibrium in finite time, as can also be seen from (4.4). The ARLs in this case, however, appear to be much shorter than with constant delays.

4.5 Conclusion

A CUSUM chart incorporating iid delays in continuous time has been proposed and studied through simulations. A proportional density assumption is used to describe the relationship between the failure distributions at the population level and that at the current facility of interest. The fact that the system of homogeneous Poisson arrivals followed by independent failure times and delays gives rise to a multivariate Poisson process is a useful connection. It is easy to see that this idea may be generalized to cases when there is a relay of events following independently after a stream of homogeneous Poisson arrivals.

In organ transplantation, the reports of failures or alive status usually come in at periodic intervals scheduled from the date of transplant. So the delay times are not

iid anymore and they will depend on the failure time as well. This also means that the delay distribution will differ by institution. It is of interest to develop a CUSUM that incorporates this kind of a delay mechanism as well.

CHAPTER V

Summary and future directions

In this chapter, we shall summarize the dissertation and note possible directions for future investigation.

5.1 Pseudo observations

An investigation of the use of pseudo observations for point and interval estimation in a multi-state event history model with censoring has been implemented. Model assumptions and simulation parameters have been selected along the lines of Andersen et al. (2003). This is also compared to point estimation based on a Cox model (Cox (1972)) and interval estimation using the bootstrap and the comparison assessed via simulations. Results from simulation studies suggest that bootstrap methods do quite well in terms of coverage probabilities, for estimating parameters in the logistic model. Particularly, when applied to the Cox model estimates, both the coverage probabilities and the average interval widths seem to be very satisfactory. The estimation procedure based on pseudo observations, however, presents difficulty in implementation and yields wider intervals. Through a single time point model, it becomes clear that truncation of the pseudo values may lead to heavy biases. It is also clear, theoretically, from a simple two-sample survival model example, that estimates based on pseudo observations are inconsistent when the censoring is

covariate-dependent.

It is interesting and useful to be able to address the numerical issues involved in solving the estimating equation for pseudo observations and develop theoretical results for the estimates based on pseudo observations for models involving censored data. In most models considered so far, the occupancy probabilities are complex non-linear functions of covariates and attempts to construct a linear relationship could lead to very crude approximations. This error in approximation may partially explain why the pseudo observations approach is not working very well. In this dissertation, we have fit a logistic model to the pseudo observations. As noted before, some of the pseudo observations are outside the range of values permissible under the model and it is not clear what a reasonable model would be. The fact that the pseudo observations need not lie in the same range as the original parameter to be estimated complicates the problem. When the censoring is covariate-dependent, pseudo observations are not able to correctly estimate the model parameters conditional on the covariate, which is essential in a regression problem.

Finally, note that pseudo observations arise from the jackknife. Previous work in this area (see Hinkley (1977), Simonoff and Tsai (1986), Wu (1986) among many others) discuss applications where the pseudo observations are generated from the model whose parameters are to be estimated. In the applications discussed in this article, however, the pseudo observations are defined prior to fitting the regression model, substituted as the response in the regression and the regression parameter estimated. So it is not clear how one may appeal to results on the jackknife based on earlier work and more research is required in this area. It is useful to investigate how these methods perform when pseudo observations are generated from the model rather than the estimate.

5.2 CUSUM procedures

A CUSUM procedure has been developed and implemented in continuous time and simulation studies demonstrate that it leads to quicker detection or signal times as compared to certain discrete versions commonly used in the health-care area. The method of likelihood ratio scoring may be applied to several settings that involve monitoring outcomes. In this dissertation, we have used a binary outcome and a Poisson approximation. However, other models may also be used to form the likelihood ratio at each time point and then define a CUSUM process in a similar fashion. Theoretical approximations to the average time to signal may also be obtained similarly by exploiting the renewal nature of the CUSUM process and employing Wald's identity.

Several other approaches have also been suggested in the literature (Dvoretzky et al. (1953), Ewan and Kemp (1960), Woodall (1983), Zacks (2004)) for deriving approximations to the ARL. Methods using integral equations have been discussed in Goel and Wu (1971). Approximations using Markov chain methods have been discussed in Brook and Evans (1972) and Steiner et al. (2000). The theoretical approximations to the ARL derived here agree with the actual simulated numbers quite well when the underlying random walk for the CUSUM has a positive drift. It is of some interest to investigate if improvements on the current approximations are possible for the case of a negative drift using some of the above methods. The theoretical approximations to the ARL have been obtained under the assumption that the random walk process is in equilibrium which constitutes monitoring the CUSUM only after a certain transient period has elapsed. Simulation studies have shown that monitoring the CUSUM from startup ($t = 0$) yields different ARLs and

it is of some interest to derive approximate theoretical results for the ARL in that case. It is also of interest to obtain approximations in the case of nonhomogeneous arrivals.

For building a CUSUM incorporating reporting delays, we have assumed that the delay times are iid. However, in reality the times of report are often fixed to occur at anniversaries with respect to the arrival times. It is of interest to build a CUSUM procedure which adapts closely, to some degree, to this situation.

From simulation studies, it was seen that for small facilities performing less than 10 transplants per year on average, the tradeoff between Type-I error and power becomes more apparent. If it is not a lengthy and costly process to review a small facility, then a little higher Type-I error may be tolerated to buy more power which may be crucial to detect performances consistent with the alternative. Relaxing the level of Type-I error or increasing the follow-up period may be possible options for getting more power, but more research is needed here.

In this dissertation, we have simulated the CUSUMs under the assumption that the performance status of an institution remains the same throughout time (see (3.4), for example). This approach has the advantage that it becomes easier to define a false alarm rate based on whether the institution is in-control or out-of-control. In actual practice, however, the state of a facility may be transitory and the postulation of a change-point may be more interesting. The CUSUM may be formulated as a change-point detection problem (see, for example, Lorden (1971), Moustakides (2004)) and it is interesting to simulate under these conditions. It is, however, not clear how one may define false alarm or misclassification rates in this scenario.

For the assessment of false alarm rates in Chapter III, one needs to know the true status of an institution. While it is easy to do this in a simulation setting, it is

not clear how to do this for real data. It may be useful to design the CUSUM first on a portion of the data with identifiers for in- and out-of-control status and then implement it on the rest of the data. Conducting the CUSUM in two phases has also been suggested in Woodall et al. (2006).

For a CUSUM chart to signal, it should essentially attempt to capture the slope of the chart and so failures in quick succession are more crucial than the absolute number of failures. As studied in this dissertation, a CUSUM having fixed boundaries is bound to signal eventually without regard to this feature occurring or not. It is interesting to develop signaling rules for the CUSUM based on the V-Mask (Barnard (1959)) procedure. This is a two-sided procedure involving drawing a cone with vertex at a fixed horizontal distance to the right from the current CUSUM value and arms diverging towards the origin. The CUSUM is said to have signaled at the point where either arm crosses the CUSUM curve and not signaled if the curve is completely within the cone. This takes into account the increasing variation of the chart with time and signals when the slope of the chart registers abrupt changes.

In so far, we have discussed the implementation of CUSUM procedures designed to detect outcomes which are worse than the average. In the interest of quality improvement, it may be desirable to have a CUSUM for identifying performance changes leading to better outcomes. It is possible to design a one-sided CUSUM with an alternative in the other direction and the CUSUM would stay below zero with downward trends implying worse performance. One can also use two-sided CUSUM charts as discussed in Page (1954) to detect performance changes in either direction.

Finally, it is important to remember that CUSUM signaling does not necessarily prove that a clinically important decline or improvement in clinical quality has

occurred. Rather, the signal suggests that closer examination by the quality improvement team may be required. Although the CUSUM has been set up as a test of hypothesis problem, this is an important philosophical distinction from the usual accept or reject outcome of a hypothesis test.

APPENDICES

APPENDIX A

Computation for censoring in the illness-death model

The probability that an individual with covariate Z is censored in the illness-death model is,

$$P_{1C}(Z) = \int_0^\infty \int_u^\infty P_{11}(0, u-|Z) d\Lambda_{12}(u) P_{22}(u+, v-|Z) d\Lambda_{2C}(v) + \int_0^\infty P_{11}(0, u-|Z) d\Lambda_{1C}(u)$$

Under assumptions of time-constant intensities, this simplifies to,

$$P_{1C}(Z) = \frac{1}{\lambda_{12}e^{\gamma_{12}Z} + \lambda_{13}e^{\gamma_{13}Z} + \lambda_{1C}e^{\gamma_{1C}Z}} \left[\lambda_{1C}e^{\gamma_{1C}Z} + \frac{\lambda_{12}e^{\gamma_{12}Z}\lambda_{2C}e^{\gamma_{2C}Z}}{\lambda_{23}e^{\gamma_{23}Z} + \lambda_{2C}e^{\gamma_{2C}Z}}\right]$$

The unconditional probability of being censored with $Z = 0$ or 1 is,

$$\alpha = \pi P_{1C}(1) + (1-\pi)P_{1C}(0)$$

APPENDIX B

Deriving the limit of $E(\tau_h)$ as $\eta \to 0$

We have,

$$E(\tau_h) = \frac{h}{\eta} - \frac{e^{-\mu}(e^{\theta}-1)}{\eta}\left(\frac{1-e^{-\omega_0 h}}{1-e^{-\omega_0\theta}}\right)$$

We can take the limit of the above expression as $\eta \to 0$. Consider first, the function $h(\omega) = \log f^*(\omega)$. We want to look for a zero of $h(\omega)$. We have, ignoring constants,

$$\begin{aligned} h(\omega) &= \omega(e^{\theta}-1) + e^{\mu}(e^{-\omega\theta}-1) \\ &= \omega(e^{\theta}-1) - e^{\mu}\left(\omega\theta - \frac{\omega^2\theta^2}{2} + \ldots\right) \\ &= -\frac{\eta}{\gamma}\omega + \frac{\omega^2\theta^2 e^{\mu}}{2} + \ldots \end{aligned}$$

so that an approximate non-trivial zero may be obtained as $\omega_0 = 2\eta/\sigma^2$, by considering only the first 2 terms and using $\sigma^2 = \theta^2 e^{\mu}\gamma$. This approximation will be good if we can ignore the higher order terms and that can be done if $\omega \to 0$. Observe that, when $\eta = 0$, $h(\omega) = e^{\mu}(\omega\theta - 1 + e^{-\omega\theta})$ and so $\omega_0\theta = 1 - e^{-\omega_0\theta}$ is an exact relation for every zero ω_0 of h. Also, one can see that $\omega_0 = 0$ is the only zero in this case.

Combining all of these, and noting that $\theta e^{\mu} = e^{\theta} - 1$ when $\eta = 0$, we get,

$$\begin{aligned}
\lim_{\eta \to 0} E(\tau_h) &= \lim_{\eta \to 0} \left[\frac{h}{\eta} - \frac{e^{-\mu}\theta e^{\mu}}{\eta} \frac{\omega_0 h - \omega_0^2 h^2/2}{\omega_0 \theta} \right] \\
&= \lim_{\eta \to 0} \left[\frac{h}{\eta} - \frac{\omega_0 h - \omega_0^2 h^2/2}{\omega_0 \eta} \right] \\
&= \lim_{\eta \to 0} \frac{\omega_0 h^2}{2\eta} \\
&= \lim_{\eta \to 0} \frac{2\eta h^2}{2\eta \sigma^2} \\
&= \frac{h^2}{\sigma^2}
\end{aligned}$$

which is the formula for $E(\tau_h)$ (before substituting $\sigma^2 = \theta^2 e^{\mu} \gamma$) under $\eta = 0$.

BIBLIOGRAPHY

BIBLIOGRAPHY

Aalen, O. O. and Johansen, S. (1978), An empirical transition matrix for for nonhomogeneous Markov chains based on censored observations., *Scandinavian Journal of Statistics* **5**, 141–150

Andersen, P. K., Borgan, O., Gill, R. D., and Keiding, N. (1993), *Statistical Models Based On Counting Processes*, Springer, New York, USA

Andersen, P. K., Hansen, M. G., and Klein, J. P. (2004), Regression analysis of restricted mean life using pseudo observations., *Lifetime Data Analysis* **10**, 335–350

Andersen, P. K. and Klein, J. P. (2007), Regression analysis for multistate models based on a pseudo-value approach, with applications to bone marrow transplantation studies., *Scandinavian Journal of Statistics* **34**, 3–16

Andersen, P. K., Klein, J. P., and Rosthoj, S. (2003), Generalised linear models for correlated pseudo-observations, with applications to multi-state models., *Biometrika* **90(1)**, 15–27

Andrei, A. and Murray, S. (2007), Regression models for the mean of the quality-of-life-adjusted restricted survival time using pseudo-observations., *Biometrics* **63(2)**, 398–404

Axelrod, D. A., Guidinger, M. K., Metzger, R. A., Wiesner, R. H., Webb, R. L., and Merion, R. M. (2006), Transplant center quality assessment using a continuously updatable, risk-adjusted technique (CUSUM)., *American Journal of Transplantation* **6**, 313–323

Aylin, P., Best, N., Bottle, A., and Marshall, C. (2003), Following Shipman: A pilot system for monitoring mortality rates in primary care., *The Lancet* **362**, 485–491

Barnard, G. A. (1959), Control charts and stochastic processes (with discussion and rejoinder)., *JOurnal of the Royal Statistical Society B* **21(2)**, 239–271

Benjamini, Y. and Hochberg, Y. (1995), Controlling the false discovery rate: a practical and powerful approach to multiple testing., *Journal of the Royal Statistical Society B* **57(1)**, 289–300

Brook, D. and Evans, D. A. (1972), An approach to the probability distribution of CUSUM run length., *Biometrika* **59(3)**, 539–549

Cox, D. R. (1972), Regression models and life tables (with discussion), *Journal of the Royal Statistical Society B* **34**, 187–220

Cox, D. R. and Miller, H. D. (1965), *The Theory of Stochastic Processes*, Chapman and Hall, London, UK

Dickinson, D. M., Shearon, T. H., O'Keefe, J., Wong, H. H., Berg, C. L., Rosendale, J. D., Delmonico, F. L., Webb, R. L., and Wolfe, R. A. (2006), SRTR center-specific reporting tools: posttransplant outcomes., *American Journal of Transplantation* **6(Part 2)**, 1198–1211

Dvoretzky, A., Kiefer, J., and Wolfowitz, J. (1953), Sequential decision problems for processes with continuous time parameter. Testing hypothesis., *Annals of Mathematical Statistics* **24(2)**, 254–264

Efron, B. (1977), Efficiency of Cox's likelihood function for censored data., *Journal of the American Statistical Association* **72**, 557–565

Efron, B. and Tibshirani, R. J. (1993), *An Introduction to the Bootstrap.*, Chapman & Hall, London

Ewan, W. D. and Kemp, K. W. (1960), Sampling inspection of continuous processes with no autocorrelation between successive results., *Biometrika* **47**, 363–380

Gan, F. F. (1992), Exact run length distributions for one-sided exponential CUSUM schemes., *Statistica Sinica* **2**, 297–312

Goel, A. L. and Wu, S. M. (1971), Determination of ARL and a contour nomogram for CUSUM charts to control normal mean., *Technometrics* pp 221–230

Grigg, O. and Spiegelhalter, D. (2005), Random-effects CUSUMs to monitor hospital mortality., in *Statistical Solutions to Modern Problems: Proceedings of the 20th International Workshop on Statistical Modeling, Sydney*, pp 239–246, University of Western Sydney Press

Grigg, O. and Spiegelhalter, D. (2006), *The null steady-state distribution of the CUSUM statistic.*, Technical report, MRC Biostatistics Unit, Institute of Public Health, Robinson Way, Cambridge UK

Hinkley, D. V. (1977), Jackknifing in unbalanced situations., *Technometrics* **19(3)**, 285–292

Kalbfleisch, J. D. and Lawless, J. F. (1985), The analysis of panel data under a Markov assumption., *Journal of the American Statistical Association* **80(392)**, 863–871

Kalbfleisch, J. D. and Prentice, R. L. (2002), *The Statistical Analysis of Failure Time Data*, Wiley-InterScience, New York, NY, USA

Khan, R. A. (1978), Wald's approximations to the average run length in CUSUM procedures., *Journal of Statistical Planning and Inference* **2**, 63–77

Klein, J., Logan, B., Harhoff, M., and Andersen, P. (2007), Analyzing survival curves at a fixed point in time, *Statistics in Medicine* **26**, 4505–4519

Klein, J. P. and Andersen, P. K. (2005), Regression modeling of competing risks data based on pseudovalues of the cumulative incidence function., *Biometrics* **61**, 223–229

Liang, K. Y. and Zeger, S. L. (1986), Longitudinal data analysis using generalized linear models., *Biometrika* **73**, 13–22

Lorden, G. (1971), Procedures for reacting to a change in distribution., *Annals of Mathematical Statistics* **42**, 1897–1908

Marshall, C., Best, N., Bottle, A., and Aylin, P. (2004), Statistical issues in the prospective monitoring of health outcomes across multiple units., *Journal of the Royal Statistical Society A* **167(3)**, 541–559

Moustakides, G. V. (2004), Optimality of the CUSUM procedure in continuous time., *The Annals of Statistics* **32(1)**, 302–315

Novick, R. J., Fox, S. A., Stitt, L., Swinamer, S., Lehnhardt, K., Rayman, R., and Boyd, W. (2001), Cumulative sum failure analysis of a policy change from on-pump to off-pump coronary artery bypass grafting., *Annals of Thoracic Surgery* **72**, 1016–1021

Page, E. S. (1954), Continuous inspection schemes., *Biometrika* **41**, 100–114

Poloniecki, J., Valencia, O., and Littlejohns, P. (1998), Cumulative risk adjusted mortality chart for detecting changes in death rate., *British Medical Journal* **316**, 1697–1700

Reynolds, M. R. (1975), Approximations to the average run length in cumulative sum control charts., *Technometrics* **17(1)**, 65–71

Shewhart, W. A. (1931), *Economic Control of Quality of Manufactured Product*, Van Nostrand, New York, USA

Siegmund, D. (1985), *Sequential Analysis Tests and Confidence Intervals.*, Springer-Verlag, New York

Simonoff, J. S. and Tsai, C. L. (1986), Jackknife-based estimators and confidence regions in nonlinear regressions., *Technometrics* **28**, 103–112

Steiner, S. H., Cook, R. J., and Farewell, V. T. (2001), Risk-adjusted monitoring of binary surgical outcomes., *Medical Decision Making* **21**, 163–169

Steiner, S. H., Cook, R. J., Farewell, V. T., and Treasure, T. (2000), Monitoring surgical performance using risk-adjusted cumulative sum charts., *Biostatistics* **1**, 441–452

Storey, J. D. (2002), A direct approach to false discovery rates., *Journal of the Royal Statistical Society B* **64**, 479–498

Therneau, T. M. and Grambsch, P. (2000), *Modeling Survival Data: Extending the Cox Model*, Springer-Verlag, New York, NY, USA

Van Dobben de Bruyn, C. S. (1968), *Cumulative Sum Tests: Theory and Practice*, Hafner Publishing Company, New York, USA

Vardeman, S. and Ray, D. (1985), Average run lengths for CUSUM schemes when observations are exponentially distributed., *Technometrics* **27(2)**, 145–150

Wald, A. (1947), *Sequential Analysis*, John Wiley and Sons, New York, USA

Woodall, W. H. (1983), The distribution of the run length of one-sided CUSUM procedures for continuous random variables., *Technometrics* **25(3)**, 295–301

Woodall, W. H., Mohammed, M. A., Lucas, J. M., Watkins, R., Steiner, S. H., Benneyan, J. C., Grigg, O. A., Spiegelhalter, D. J., and H., B. (2006), The use of control charts in health-care and public-health surveillance (with discussion and rejoinder)., *Journal of Quality Technology* **38(2)**, 89–134

Wu, C. F. J. (1986), Jackknife, bootstrap and other resampling methods in regression analysis., *The Annals of Statistics* **14(4)**, 1261–1295

Zacks, S. (2004), Exact determination of the run length distribution of a one-sided CUSUM procedure applied on an ordinary Poisson process., *Sequential Analysis* **23**, 159–178