Applications of Control Charts in Medicine and Epidemiology

Landon H. Sego

Dissertation submitted to the faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in the

Department of Statistics

William H. Woodall
Marion R. Reynolds, Jr.
Jeffrey B. Birch
Dan J. Spitzner
G. Geoffrey Vining

5 April 2006
Blacksburg, Virginia

KEY WORDS: monitoring, Sets, CUSCORE, CUSUM, steady-state, control chart, risk-adjusted, survival time.

© 2006, Landon H. Sego
ALL RIGHTS RESERVED
Applications of Control Charts in Medicine and Epidemiology

Landon H. Sego

Abstract

We consider two applications of control charts in health care. The first involves the comparison of four methods designed to detect an increase in the incidence rate of a rare health event, such as a congenital malformation. A number of methods have been proposed: among these are the Sets method, two modifications of the Sets method, and the CUSUM method based on the Poisson distribution. Many of the previously published comparisons of these methods used unrealistic assumptions or ignored implicit assumptions which led to misleading conclusions. We consider the situation where data are observed as a sequence of Bernoulli trials and propose the Bernoulli CUSUM chart as a desirable method for the surveillance of rare health events. We compare the steady-state average run length performance of the Sets methods and its modifications to the Bernoulli CUSUM chart under a wide variety of circumstances. Except in a very few instances we find that the Bernoulli CUSUM chart performs better than the Sets method and its modifications for the extensive number of cases considered.

The second application area involves monitoring clinical outcomes, which requires accounting for the fact that each patient has a different risk of death prior to undergoing a health care procedure. We propose a risk-adjusted survival time CUSUM chart (RAST CUSUM) for monitoring clinical outcomes where the primary endpoint is a continuous, time-to-event variable that is right censored. Risk adjustment is accomplished using accelerated failure time regression models. We compare the average run length performance of the RAST CUSUM chart to the risk-adjusted Bernoulli CUSUM chart, using data from cardiac surgeries to motivate the details of the comparison. In order to make the comparisons between the two charts as fair as possible, the RAST CUSUM chart is based on the assumption that the survival times follow a log-logistic distribution. The comparisons show that the RAST CUSUM chart is more efficient at detecting deterioration in the quality of a clinical procedure than the risk-adjusted Bernoulli CUSUM chart, especially when the fraction of censored observations is not too high. We also present a RAST CUSUM chart based on the Weibull distribution as an alternative to the log-logistic RAST CUSUM chart. We address details regarding the implementation of a prospective monitoring scheme using the RAST CUSUM chart.
Acknowledgements

The undertaking of a work as extensive as this could never be accomplished without the contribution and support of many individuals. While there are many I could mention here, there are a number of persons to whom I am particularly indebted. I wish to thank Bill Woodall and Marion Reynolds Jr. for their faithful service as my advisors. Their combined insights, their complimentary strengths, and their careful readings of many manuscripts have greatly improved the quality of my research. Their influence during the final chapters of my formal education has shaped the scientist and statistician I have aspired to become. I am grateful to Mike Box, Jeff Norris, and Mike Rebich for their excellent support of the computing facilities here in the Department of Statistics at Virginia Tech—for without that support this undertaking would not have been possible.

I am also grateful to my father, Lane Sego, and my father-in-law, Win Duersch, whose exemplary encouragement and support shone brightly during the most difficult moments of my journey. There really aren’t words that can express the depth of gratitude I feel for my wife, Heidi. Her strength and companionship have never faltered. She and our three (almost four) wonderful children have been my greatest joy and consolation—not to mention my principal motivation.

In addition, I would like to thank Dr. Tom Treasure at Guy’s Hospital, St. Thomas Street, London, Greater London, SE1 9RT UK, for access to the cardiac surgery data discussed in Part II. My thanks to Dr. Jeffrey Roylance for a most helpful discussion regarding medical conditions associated with high mortality rates. Last of all, a portion of the research discussed in Part I was supported by NSF Grant DMI-0354859.

— Landon Sego
Contents

List of Figures viii

List of Tables x

1 General Introduction 1

I Continuously monitoring a small incidence rate 3

2 Introduction 4

3 Description of the methods 9

3.1 Sets method . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 10
3.2 SHDA method . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 11
3.3 CUSCORE method . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 13
3.4 Bernoulli CUSUM . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 14

4 Evaluation of chart performance 16

4.1 Considerations in assessing chart performance . . . . . . . . . . . . . . . . . . . . . . . . . 16
4.1.1 Defining the average run length . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 17
4.1.2 Steady-state average run length . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 17
4.1.3 Head-start features . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 18
4.2 Sets method . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 19
4.3 SHDA method . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 23
4.4 CUSCORE method . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 26
4.5 Derivation of $\beta = P(Z < t)$ for Sets, SHDA, and CUSCORE methods . . . . . . . . 27
### 4.6 Derivation of SSANB for Sets, SHDA, and CUSCORE Methods

4.7 Bernoulli CUSUM

4.8 Revised explicit solutions of the ANB for the CUSUM

4.9 Algorithm for the simulation of the CUSUM SSANB

4.10 Weighted average of the ARL’s for an arbitrary shift time

4.11 Measuring SSANB performance across a range of possible shifts

### 5 Determining the design parameters

5.1 Optimality criteria

5.2 Sets method

5.3 SHDA method

5.4 CUSCORE method

5.5 Bernoulli CUSUM

### 6 Comparison of methods

6.1 Design of the study

6.2 Comparison of methods using SSANB

6.3 Comparison of the methods using the area under the SSANB curves

6.4 Effects of $p_0$, $\gamma_1$, and $m_0$ on chart performance

### 7 Discussion

7.1 Further considerations in assessing chart performance

7.1.1 Measuring performance using the initial state ANB

7.1.2 Measuring performance using the SSANB

7.1.3 Measuring performance for an arbitrary shift time

7.2 Criteria for choosing the design parameters

7.3 Regarding potential bias against the sets based methods

7.3.1 Effect of the optimality criteria used to select design parameters

7.3.2 Regarding the head-start feature of the sets based methods

7.4 Alternative CUSUM approaches for monitoring a small incidence rate

7.5 Guidelines for practitioners
8 Conclusions

8.1 In search of the most equitable comparison ........................................................................ 84
8.2 Summary of results ........................................................................................................... 86
8.3 Final remarks ................................................................................................................... 87

II Risk-adjusted monitoring of clinical outcomes .................................................................. 88

9 Introduction ......................................................................................................................... 89

10 Description of the risk-adjusted CUSUM charts ................................................................. 92

10.1 Risk-adjustment ............................................................................................................. 92
10.2 The risk-adjusted Bernoulli CUSUM chart ................................................................... 94
10.3 The risk-adjusted survival time CUSUM chart ................................................................. 97
10.3.1 Constructing the risk-adjusted likelihood ratio ............................................................. 97
10.3.2 Modeling the survival distribution for each patient ....................................................... 99
10.3.3 Specifying parameter shifts in the RAST CUSUM chart ............................................. 101
10.4 The log-logistic RAST CUSUM chart .......................................................................... 103
10.5 The Weibull RAST CUSUM chart ................................................................................ 104

11 Determining the average run length .................................................................................. 106

11.1 General considerations in using the ARL as a performance metric ................................. 106
11.2 Markov chain approach for calculating the ARL ............................................................. 107
11.2.1 Transition probabilities for the RA Bernoulli CUSUM chart ....................................... 109
11.2.2 Transition probabilities for the log-logistic RAST CUSUM chart ............................... 111

12 Comparisons ..................................................................................................................... 114

12.1 Cardiac surgery example ............................................................................................... 114
12.2 Logistic regression versus the accelerated failure time regression model ....................... 115
12.3 Interpreting the accelerated failure time regression model ............................................ 118
12.4 Comparisons of the RA Bernoulli CUSUM and the RAST CUSUM charts ..................... 119

13 Discussion ........................................................................................................................ 124

13.1 Dynamics of the risk-adjusted CUSUM charts ............................................................... 124
13.2 Prospective monitoring using RAST CUSUM charts .................................................... 126
List of Figures

3.1 Grass-plot for the Sets method with $n_s = 3$ and $t_s = 1500$. Each slanted line represents a set. 11

3.2 Grass-plot for the SHDA method with $n_t = 2$, $b_t = 4$, and $t_t = 1500$. Each slanted line represents a set. Note that a flag is raised just prior to monitoring. 12

3.3 Example of the CUSCORE plot for $n_c = 4$ and $t_c = 1500$. Note that $C_i$ increases when $X_i < t_c$ and decreases (or stays at 0) when $X_i \geq t_c$. 13

3.4 Example of the CUSUM chart for $\delta = 0.15$ and $h = 2.7$. Note that $B_j$ decreases by $\delta$ for each healthy birth and increases by $1 - \delta$ when a malformation is observed. 15

4.1 Ratio of Approx $ANM$ to Exact $ANM$ for the Sets method when the chart is designed for shifts of size $\gamma_1 = 2$ and target in-control $ANM$ is 50. 20

5.1 Finding optimal $(n_s, t_s)$ for the Sets method when $p_0 = 0.001$, $\gamma_1 = 2$, and $m_0 = 50$. The solid points in the plots indicate the optimal values of $n_s$, $t_s$, $ANM_s(p_1)$, and the value of $ANM_s(p_0)$ that was achieved for $n_s = 5$ and $t_s = 747$. 44

5.2 Surface of $W_t$ for various combinations of $n_t$ and $b_t$ for $p_0 = 0.001$ and $m_0 = 200$ ($b_0 = 200000$) to detect a shift of size $\gamma_1 = 3$. In this case the minimum $W_t$ occurs when $n_t = 3$ and $b_t = 3$. 45

6.1 Forty cases of $p_0$, $\gamma_1$, and $m_0$ that were used to compare the four methods with respect to $SSANB$ performance. 51

6.2 Comparing $SSANB$ performance across a range of hypothetical shifts for cases 12 and 13. 56

6.3 Comparing $SSANB$ performance across a range of hypothetical shifts for cases 28 and 40. Since the $SSANB_b$ is simulated for case 40, the width of the solid line for the CUSUM is given by $SSANB_b(\gamma) \pm 3.63 \times SE$. 57

6.4 Plots of $y_1$ versus $p_0$, $\gamma_1$, and $m_0$ with fitted values superimposed. In each plot, the fitted regression line was calculated by allowing the predictor variable of interest to vary while the other two predictor variables were fixed at their respective means. 61

6.5 Plots of $y_2$ versus $p_0$, $\gamma_1$, and $m_0$ with fitted values superimposed. In each plot, the fitted regression line was calculated by allowing the predictor variable of interest to vary while the other two predictor variables were fixed at their respective means. 62
7.1 Convergence of \(WANB\) to \(SSANB\) for case 11 \((p_0 = 0.007, \gamma_1 = 3.75, m_0 = 100)\). For the top horizontal axis of the Sets, SHDA, and CUSCORE methods, the number of births was approximated as \(1/p_0\) births per incident. The horizontal dashed lines denote the bands inside which convergence was declared to have occurred. The vertical dotted lines mark the step when convergence occurred.

7.2 Histograms of the number of births as a fraction of \(ANB(p_0)\) at which convergence to \(SSANB(\gamma_1a)\) occurs.

10.1 Survival function of the Weibull and log-logistic distributions for different values of the acceleration factor, \(a\), and various combinations of shape parameter \(\alpha\) and scale parameter \(\lambda\).

12.1 Histogram of the Parsonnet score for cardiac patients who underwent surgery in 1992 and 1993.

12.2 Each curve represents the 30-day predicted mortality of the AFT model subtracted from the 30-day predicted mortality of the logistic regression model that were fit to one of the 150 simulated data sets.

12.3 Visualizing the predicted survival function of the log-logistic AFT regression model for various values of the Parsonnet score \((u)\) versus the non risk-adjusted Kaplan-Meier estimate of the survival function.

12.4 Ratio of the \(ARL\) for the log-logistic RAST CUSUM chart to the \(ARL\) of the RA Bernoulli CUSUM chart for the four cases, each with a different level of baseline censoring.

12.5 Average number of patients not exposed to higher risk as a result of using the log-logistic RAST CUSUM chart instead of the RA Bernoulli CUSUM chart. To give better resolution to case 1 (97.5% baseline censoring), the vertical axis is presented in the log scale.

13.1 Weibull and log-logistic RAST CUSUM scores for the training data set of the cardiac surgery example. The lines represent the scores when an actual death time is observed. The solid circles represent the value of the score when the survival time is censored at 30 days. The corresponding Parsonnet score is indicated to the right of the circles.

A.1 \(SSANB\) profiles for cases 1 to 6.

A.2 \(SSANB\) profiles for cases 7 to 12.

A.3 \(SSANB\) profiles for cases 13 to 18.

A.4 \(SSANB\) profiles for cases 19 to 24.

A.5 \(SSANB\) profiles for cases 25 to 30.

A.6 \(SSANB\) profiles for cases 31 to 36. To reflect the variability in the simulated values of \(SSANB_h(\gamma)\) for the CUSUM, in cases 34 and 36, the thickness of the solid line for the CUSUM is equal to width of the 95% simultaneous Bonferroni confidence interval.

A.7 \(SSANB\) profiles for cases 37 to 40. To reflect the variability in the simulated values of \(SSANB_h(\gamma)\) for the CUSUM, in cases 37 to 40, the thickness of the solid line for the CUSUM is equal to width of the 95% simultaneous Bonferroni confidence interval.
List of Tables

4.1 States for the SHDA method with $n_t = 3$ and $b_t = 5$. In the third and fourth columns, 1 represents the event $\{X_i < t_s\}$ and 0 represents the event $\{X_i \geq t_s\}$. ........................................... 24

4.2 Algorithmic listing of states for the SHDA method (first read down the columns and then left to right). ......................................................................................... 25

6.1 Cases of $p_0$, $\gamma_1$, and $m_0$ with optimal design parameters ($w$ is the number of transient states in the Markov chain for the CUSUM). ............................................................... 52

6.2 Values of $ANB(p_0)$ that were achieved for the 40 cases of $p_0$, $\gamma_1$, and $m_0$. .......................... 53

6.3 Values of $SSANB(\gamma_1a)$ for the 40 cases of $p_0$, $\gamma_1$, and $m_0$. The 95% simultaneous confidence interval is $SSANB_b(\gamma_1a)^0 = 3.63 \times SE$ for the six cases when the CUSUM was simulated. ... 55

6.4 Thirteen instances when the CUSCORE had better $SSANB$ performance than the CUSUM. The ratio is $SSANB_b(\gamma)/SSANB_c(\gamma)$. ................................. 57

6.5 Values of $SSANB(\gamma)$ for case 12 ($p_0 = 0.006$, $\gamma_1 = 1.25$, $m_0 = 400$), case 13 ($p_0 = 0.006$, $\gamma_1 = 2.25$, $m_0 = 350$), case 28 ($p_0 = 0.0009$, $\gamma_1 = 3.5$, $m_0 = 850$), and case 40 ($p_0 = 0.0002$, $\gamma_1 = 5.75$, $m_0 = 350$) for various hypothetical shifts of size $\gamma$. The 95% simultaneous confidence interval is $SSANB_b(\gamma)^0 = 3.63 \times SE$ for case 40. ........................................ 58

6.6 Approximate areas under the log($SSANB$) curves for the 40 cases ($\approx \int_{1.25}^{7.75} \log(SSANB(\gamma))d\gamma$). For the simulated cases indicated by the asterisk, the area was computed under the upper bound of the 95% simultaneous confidence interval ($\approx \int_{1.25}^{7.75} \log(SSANB_b(\gamma) + 3.63 SE)d\gamma$). 59

6.7 Summary of Table 6.6, indicating the number of cases for each method where $\int_{1.25}^{7.75} \log(SSANB(\gamma))d\gamma$ was least among the four methods (“Best”), second least among the four methods (“2nd best”), etc. ......................................................... 60

6.8 Fitted regression models to assess the effect of $p_0$, $\gamma_1a$, and $m_0$ on the responses $y_1$ and $y_2$. 61

7.1 Values of $ANB(\gamma_1a_p_0)$ of the four methods evaluated for each of the 40 cases. For the six simulated cases, the 95% simultaneous confidence interval is given by $ANB(\gamma_1a_p_0) \pm 3.63 SE$. The Ratio column is the CUSUM $ANB(\gamma_1a_p_0)$ divided by the lowest $ANB(\gamma_1a_p_0)$ among the four methods for the given case. ........................................... 65

7.2 Summary of Table 7.1, indicating the number of cases for each method where the $ANB(\gamma_1a_p_0)$ was least among the four methods (“Best”), second least among the four methods (“2nd best”), etc. ............................................................................. 66

7.3 Number of births as a fraction of $ANB(p_0)$ at which $WANB(\gamma_1a)/SSANB(\gamma_1a)$ falls and remains within the interval $(0.999, 1.001)$. Eleven cases for the CUSUM were omitted because they were not computationally feasible. ........................................... 78
7.4 Thirteen instances when the $SSANB(\gamma)$ performance of the CUSCORE or the SHDA methods were better than the CUSUM in the smaller study with 23 cases. 

7.5 Values of $SSANB(3.5)$ for case 26 when the methods are adjusted to remove the effect of the head-starts of the sets based methods and when they are not adjusted. 

7.6 Values of $w$ for the Bernoulli CUSUM with the corresponding actual $ANB_0(p_0)$ underneath in parentheses. Note that $h = w/m$. 

7.7 Values of $w$ for the Bernoulli CUSUM with the corresponding actual $ANB_0(p_0)$ underneath in parentheses. Note that $h = w/m$. 

10.1 Parsonnet risk factors. To calculate the Parsonnet score, add all weights that apply to the patient and the type of operation—for example, a female patient having tricuspid valve replacement has an estimated risk of $1+3=4\%$. CABG = coronary artery bypass grafting and PASP = pulmonary artery systolic pressure. The content of this table was given by Poloniecki et al. [58]. 

12.1 Parameter values for the four cases. 

13.1 Illustration of the RA Bernoulli CUSUM score statistic, $W_B^B(y_i)$, for various values of the Parsonnet score, $u_i$. 

B.1 Results from cases 1 and 2. For case 1, $h^+_L = 4.832743$ and $h^+_B = 4.798883$. For case 2, $h^+_L = 6.0731$ and $h^+_B = 6.006768$. 

B.2 Results from cases 3 and 4. For case 3, $h^+_L = 6.241462$ and $h^+_B = 6.021252$. For case 4, $h^+_L = 6.27189$ and $h^+_B = 5.5847$. 

xi
Chapter 1

General Introduction

This work is divided into two parts, each of which contains a discussion of a different problem in the areas of monitoring rare events and monitoring the quality of a clinical procedure. In Part I we discuss the continuous monitoring of a rare health event, such as a birth defect. A number of methods have been proposed for the surveillance of birth defects. Among these are the Sets method, two modifications of the Sets method, and the CUSUM method based on the Poisson distribution. Many of the previously published comparisons of these methods used unrealistic assumptions or ignored implicit assumptions which led to misleading conclusions. We consider the situation where data are observed as a sequence of Bernoulli trials and propose the Bernoulli CUSUM chart as a desirable method for the surveillance of rare health events. We compare the steady-state average run length performance of the Sets methods and its modifications to the Bernoulli CUSUM chart under a wide variety of circumstances. Except in a very few instances we find that the Bernoulli CUSUM chart performs better than the Sets method and its modifications for the extensive number of cases considered.

In Chapter 2 in Part I we give a more detailed overview of the problem of monitoring rare health events and review the pertinent literature. In Chapter 3, we present each of the four methods under consideration: the Sets method, its two modifications, and the Bernoulli CUSUM chart. In Chapter 4, we discuss the various metrics that are used to evaluate and compare the performance of the methods and we provide the pertinent derivations of these metrics. In Chapter 5 we address the optimal choice of the design parameters for each of the methods. In Chapter 6 we present the comparisons of the four methods under a wide variety of monitoring situations. In Chapter 7 we revisit several issues concerning the comparison and evaluation of surveillance methods and provide some practical suggestions for the implementation of the Bernoulli
CUSUM chart. In Chapter 8 we give our conclusions regarding the surveillance of rare health events.

In Part II we discuss the second application area of monitoring clinical outcomes, which requires accounting for the fact that each patient has a different risk of death prior to undergoing a health care procedure. Whereas in Part I we assume that the data arrive as a sequence of Bernoulli trials where the in-control probability of the event of interest is constant across all individuals, in Part II we assume that we observe a survival time (or a censored time) for each patient, and that these observations are included in the control chart in the order in which the patients underwent the procedure. In addition, the characteristics of the survival distribution are not assumed to be the same across all patients. To this end, we propose a risk-adjusted survival time CUSUM chart (RAST CUSUM) for monitoring clinical outcomes that are characterized by a continuous, time-to-event variable that is right censored. Risk adjustment is accomplished using accelerated failure time regression models. We compare the average run length performance of the RAST CUSUM chart to the risk-adjusted (RA) Bernoulli CUSUM chart, using the outcomes from cardiac surgery to motivate the details of the comparison. In order to make the comparisons between the two charts as fair as possible, the RAST CUSUM method is based on the assumption that the survival times follow a log-logistic distribution. The comparisons show that the RAST CUSUM chart is more efficient at detecting deterioration in the quality of a clinical procedure than the RA Bernoulli CUSUM chart, especially when the censoring fraction is not too high. We also present a RAST CUSUM chart based on the Weibull distribution as an alternative to the log-logistic RAST CUSUM chart. We address details regarding the implementation of a prospective monitoring scheme using the RAST CUSUM chart.

In Chapter 9 (the first chapter in Part II), we set forth the problem of risk-adjusted monitoring of clinical outcomes along with a brief review of the pertinent literature. In Chapter 10 we present the RA Bernoulli CUSUM chart and two varieties of the RAST CUSUM chart. In Chapter 11 we discuss approaches for calculating the average run length of the charts. Chapter 12 contains a comparison of the RA Bernoulli CUSUM chart and the RAST CUSUM chart based on the log-logistic distribution, motivated by an example of cardiac surgery outcomes in the United Kingdom. In Chapter 13 we examine the dynamics of the risk-adjusted charts. We also discuss several points regarding the implementation of a prospective monitoring scheme using the RAST CUSUM chart. Last of all, in Chapter 14, we give our conclusions regarding the risk-adjusted monitoring of clinical outcomes.

The notation and symbols used in the two Parts of this work should be considered independently of one another.
Part I

Continuously monitoring a small incidence rate
Chapter 2

Introduction

The problem of monitoring the incidence rate of an event of interest where the baseline probability of the event is small arises in industrial, medical, and epidemiological settings. One especially pertinent example is monitoring the incidence of a rare type of congenital malformation. The discussion of how to best monitor congenital malformations began in earnest in the late 1960’s [1, 2, 3]. Since then, a number of statistical methods for the surveillance (or monitoring) of an incidence rate have been proposed. (While some authors [3] distinguish between “monitoring” and “surveillance,” we use the terms interchangeably.) A comprehensive review of statistical surveillance in public health was given by Sonesson and Bock [4]. Earlier reviews which focused specifically on the continuous monitoring of malformation incidence were given by Barbujani [5] and Lie et al. [6].

In this work, we focus on methods designed to detect a sudden increase in an incidence rate. We consider the commonly discussed Sets method [7] and two noteworthy modifications: the CUSCORE method [8] and an extension of the Sets method proposed by Sitter, Hanrahan, DeMets, and Anderson [9] which we will refer to as the SHDA method. We will collectively refer to these three methods as the sets based methods. We also consider the Bernoulli cumulative sum (CUSUM) chart as a method for detecting increases in a small incidence rate. While the general CUSUM methodology was proposed by Page [10], we will use a formulation of the Bernoulli CUSUM chart that was set forth by Reynolds and Stoumbos [11, 12]. We will compare and discuss these four methods in the context of birth defect surveillance. However, the methods themselves and the results presented herein are more broadly applicable to any situation involving monitoring the incidence rate of a rare event.

The sets based methods use the “distance” between incidents as the basis for detecting changes in the
incidence rate. With regard to congenital malformations, this distance can be either the number of infants born between malformed cases or, if the birth rate is assumed to be constant, the amount of time that elapses between incidents. In the former case, the number of infants born between malformed cases is a geometric variate. Chen originally designed the Sets method using the geometric model [7]. With regard to the latter case, the sets based methods are easily adapted to monitor the time elapsed between incidents, provided we assume that the incidence of malformations follows a homogeneous Poisson process—and thus the time elapsed between incidents is distributed exponentially. This approach was used by Sitter et al. [9].

If the distribution of the number of incidents per unit time has a variance that is larger than the mean (overdispersion), the negative binomial distribution can be used to model the occurrence of malformations [1]. Gallus et al. [13] used the Sets method to examine the effect of assuming the arrival of incidents follows a Poisson process when, in fact, the process is negative binomial. They determined that “over a large range of practical instances, the Poisson assumption provides a reasonable description of the negative binomial process,” even though the misspecification does result in a higher frequency of false alarms than expected.

The Poisson assumption leads naturally to the idea of monitoring the number of events per unit of time using a CUSUM chart based on the Poisson distribution [1, 3, 15]. Many researchers have compared the Poisson CUSUM to the Sets or the CUSCORE methods [7, 16, 17, 18, 19]. Gallus et al. [17] and Radaelli [19] found that the Poisson CUSUM signals an alarm more quickly than the Sets method when there is less than a four-fold increase in the incidence rate, while the Sets method signals faster for increases that are greater than four-fold. Barbujani [16] concluded in a simulation study that the Poisson CUSUM in general has faster signal times and greater sensitivity (the probability of signaling an alarm when the incidence rate has, in fact, increased), greater specificity (the probability of not signaling an alarm when the incidence rate has not changed), and better accuracy (the probability that the decision made by using the chart is correct) than the Sets method. However, Chen [20] questioned the methodology employed in that study. Chen [18] determined that the Sets method signals an alarm more quickly than the Poisson CUSUM when the average number of incidents per year is 5 or less and when the increase in the incidence rate is small. Otherwise, the Poisson CUSUM has better performance. The Sets method has also been compared to a modified geometric CUSUM chart [14], the CUSCORE method [8, 19, 21], and the SHDA method [9, 22, 23]. These studies generally concluded that the modified geometric CUSUM, the CUSCORE, and the SHDA method appear to outperform the Sets method in most instances.

Careful examination of the published comparisons involving the sets based methods reveals four important issues that have a direct impact on the conclusions reached by these studies. First, with only
one exception [8], these comparisons were made using the initial state average run length (\textit{ARL}), which assumes that the shift in the incidence rate occurs prior to the onset of monitoring. Second, none of the comparisons involving the sets based methods acknowledge the head-start features that are implicit in these three methods. Third, in some instances, comparisons were made between charts that were designed using different optimality criteria [7, 16, 18, 14, 24]. And fourth, the performance of the methods was only compared at the shift size that the chart was specifically designed to detect and not across a range of possible shift sizes [8, 9, 14, 17, 18, 19, 24]. We briefly address these four points below.

**Point 1:** The initial state \textit{ARL} is a metric that is commonly used to compare the performance of surveillance methods and control charts. In the context of the surveillance of birth defects, the term “initial state \textit{ARL}” refers to the expected number of births (or malformations) from when monitoring begins (in the initial state of the chart) until an alarm is signaled. Comparing the performance of two or more charts using the initial state \textit{ARL} assumes that the shift takes place prior to the onset of monitoring (or just as monitoring begins)—an overly restrictive and unrealistic assumption. In the typical monitoring paradigm, one begins monitoring during a non-epidemic period with the goal of detecting an increase in the baseline incidence rate should one occur. Of course, there always exists the possibility that the shift has already occurred before monitoring begins—but one cannot ignore the arguably more probable situation in which the shift occurs sometime after monitoring has begun.

**Point 2:** A control chart (or surveillance method) is said to have a head-start when the \textit{ARL} from the time of the shift until signal is shorter if the shift occurs before monitoring begins than if the shift occurs sometime after monitoring has commenced. Hence, a chart with a head-start feature reacts more quickly if an epidemic is already underway prior to monitoring. This issue is directly related to the first point because assuming that the shift occurs prior to monitoring puts undue emphasis on the head-start feature of the sets based methods. Because the effect of the head-start diminishes as monitoring proceeds, if the shift takes place after monitoring begins, the sets based methods do not react as quickly as they would if the shift were to occur prior to monitoring. Accounting for the head-start is of particular importance when comparing a chart that has a head-start feature to one that does not. For example, the traditional Poisson CUSUM chart does not have a head-start feature. Therefore, comparing the Poisson CUSUM chart to the Sets, CUSCORE, or SHDA methods by using the initial state \textit{ARL} puts the Poisson CUSUM at a disadvantage. Likewise, the SHDA method has a head-start over the Sets method—and thus comparing these two methods using the initial state \textit{ARL} confers an unfair advantage upon the SHDA method.

**Point 3:** Typically, surveillance methods and control charts are designed to minimize their delay in
reacting to a specific shift in the incidence rate, subject to not exceeding a prespecified false alarm rate. There are a number of reasonable mathematical criteria that can be used to identify the “optimal” values of the charts’ design parameters. Comparing two or more methods that were designed to be optimal under different criteria can inadvertently favor one method over the other. Thus, careful consideration must be given to the criteria used to determine the design parameters of the methods.

**Point 4:** Before monitoring begins, many methods require that the shift size be specified in order to optimally design the chart. However, once monitoring is underway, it is obviously unlikely that a real shift would coincide with the prespecified shift size that was used to design the chart. Rather than comparing the performance of two or more surveillance methods only at the shift for which they were designed to be optimal, a more realistic comparison should include some assessment of how the charts perform across a range of possible shifts. Chen [25] proposed a method for choosing the design parameters that would be robust to a variety of shift sizes and Radaelli [26] also considered this problem but used a different (and, in our view, more reasonable) analytical approach. However, these authors did not empirically investigate their approaches by comparing different methods over a broad range of possible shifts.

In this work, we assume that the data can be observed as a sequence of independent Bernoulli trials. This, of course, is the case if every infant born at a hospital (or perhaps in a larger, well-defined monitoring region) can be examined and diagnosed sequentially. If diagnosis of every infant is not feasible, it may still be possible to monitor the Bernoulli sequence from a subset of infants that were examined—although this admittedly limits the scope of the inference we could make and increases the potential for biased results. If we measure the “distance” between incidents as the number of births between malformed cases, the Sets method and its modifications are easily employed under the Bernoulli assumption. In addition, the observation of a sequence of Bernoulli trials leads naturally to the idea of using a CUSUM chart based on the Bernoulli distribution to monitor the incidence rate.

The primary objective of Part I is to compare the performance of the sets based methods to the Bernoulli CUSUM chart. We have approached this comparison in a way that specifically addresses the four concerns we have raised about previously published comparisons. Ultimately, the goal is to compare the methods in the most equitable way possible under the least restrictive assumptions.

We do not include the Poisson CUSUM among the comparisons in this work because the Bernoulli CUSUM is more efficient than the Poisson CUSUM when the objective is to detect an increase in the incidence rate. This is due to the fact that with the Poisson CUSUM, one artificially aggregates the data by counting the number of malformations that arise during a specified monitoring interval (e.g. monthly or
quarterly)—thereby requiring that the monitoring interval be completed before it can react to an increase in the baseline incidence rate. The Bernoulli CUSUM chart, however, operates in “real-time” and reacts to an increase without such a delay. We discuss this in more detail in Section 7.4.
Chapter 3

Description of the methods

During the surveillance of congenital malformations, the incidence rate could change in any number of ways. The rate could either decrease or increase and the change could be immediate or gradual. All four of the methods considered here could be adapted to detect decreases in the incidence rate. Likewise, they each could be constructed as two-sided schemes designed to detect both increases and decreases in the incidence rate. However, we only consider the scenario of a sustained step-shift, where the incidence rate increases suddenly by a fixed amount.

To monitor the incidence rate, we consider a sequence of Bernoulli trials $V_1, V_2, V_3, \ldots$ that represent each birth, where $V = 1$ indicates the presence of the malformation and $V = 0$ indicates its absence. We assume the sequence of random variables are independent with incidence rate $p = P(V = 1)$. Under non-epidemic conditions when the incidence of malformations remains at the baseline level, the incidence rate is $p = p_0$. Under epidemic conditions, we assume the incidence rate is some unknown $p$ where $p = \gamma p_0$ for suitable values of $\gamma$ such that $0 < p < 1$ and $p \neq p_0$. Since we are concerned with increases in the incidence rate, we generally assume that $1 < \gamma < 1/p_0$. We let $\tau \geq 0$ represent the position of the birth after which the shift in $p$ takes places. Hence $P(V_j = 1) = p_0$ for $j = 1, 2, \ldots, \tau$ and $P(V_j = 1) = p$ for $j = \tau + 1, \tau + 2, \ldots$.

For each of the four methods, we design the charts to be optimal for detecting a nominal shift from $p_0$ to $p_1 = \gamma_1 p_0$, for $1 < \gamma_1 < 1/p_0$. Hence $\gamma_1$ refers to the size of the shift the chart is optimally designed to detect (optimality is formally defined in Section 5.1). A formal definition of each of the four methods is given below.
3.1 Sets method

We define a set as the group of infants that are born consecutively between those born with a particular type of congenital malformation. Let $X_i$ represent the size of the $i^{th}$ set, i.e., the number of births between (but not including) incidents $i - 1$ and $i$ for $i = 1, 2, \ldots$. The $X_1, X_2, \ldots$ form a sequence of independent geometric random variables with mass function $P(X = x) = p(1 - p)^{x}$ and take on values $\{0, 1, 2, \ldots\}$. Note that $i$ not only indexes the sequence of observed malformations but it also indexes the sets themselves. The Sets method signals an alarm if $n_s$ consecutive sets each have size less than a threshold $t_s$. (The $s$ subscript is added to clarify that the design parameters correspond to the Sets method). Assuming that the baseline incidence rate $p_0$ is known, the Sets method is characterized by the two design parameters $n_s$ and $t_s$.

It is convenient to represent the Sets method as follows. When the $i^{th}$ malformation is observed, let $S_i$ represent the number of consecutive sets that have sizes less than $t_s$, i.e.,

$$
S_0 = 0 \\
S_i = (1 + S_{i-1}I_{S_{i-1} < n_s})I_{(X_i < t_s)} \quad i = 1, 2, \ldots.
$$

(3.1)

An alarm is signaled when $S_i = n_s$. Note that equation (3.1) implies that the counter $S_i$ resets after a signal. If $n_s = 1$, the Sets method reduces to a one-sided geometric Shewhart chart with a lower control limit at $t_s$.

There are a number of ways to graphically represent the Sets method. We prefer the “grass-plot” approach used by Grigg and Farewell [27], where each set is represented as a slanted line. The line represents the string of healthy births that terminates when a malformation is observed. An example of this type of plot is shown in Figure 3.1.

The Sets method is actually a special case of a runs rule monitoring scheme discussed earlier by Page [28]. Page’s Rule IV says to take action (i.e. signal an alarm) if $n_s$ consecutive points fall between the warning and action lines or if any point falls outside the action lines. In the context of the Sets method, a point “between the warning and action lines” corresponds to the event $\{X_i < t_s\}$. The action line can be any value $a$ such that a point falling outside the action line, denoted by the event $\{X_i < a\}$, has probability zero. However, if we let $a$ assume a value in the interval $[1, t_s)$, then Page’s rule becomes the Sets method combined with a Shewhart limit, an idea later proposed by Wolter [8].
3.2 SHDA method

For the SHDA method, we define $X_i$, $n_t$, and $t_t$ just as we did for the Sets method, only we use the $t$ subscript to designate that the design parameters correspond to the SHDA method. To implement the method, if $n_t$ consecutive sets each have size less than $t_t$, then we raise a flag. When a flag is raised, if the total number of sets that have occurred since the last flag is less than or equal to a threshold $b_t$, the method signals an alarm. Assuming a known baseline incidence rate $p_0$, the SHDA method is characterized by three design parameters: $n_t$, $t_t$, and $b_t$. Figure 3.2 gives a grass-plot representation of the SHDA method and describes how an alarm is signaled when $n_t = 2$, $b_t = 4$, and $t_t = 1500$. Note that the SHDA method assumes that a flag is raised just prior to monitoring. As a result, the SHDA method begins monitoring in a state that is not furthest away from signal, thereby giving the SHDA method a head-start. We discuss this further in Section 4.1.3.

To more formally define the SHDA method, when the $i^{th}$ malformation is observed, let $G_i$ represent the number of consecutive sets that have sizes less than $t_t$. Let $T_i$ represent the total number of sets (the
number of incidents) that have occurred since the last flag. Then

\[
G_0 = 0 \\
G_i = (1 + G_{i-1} I_{\{G_{i-1} < n_t\}}) I_{\{X_i < t_t\}} \quad i = 1, 2, \ldots \\
T_0 = 0 \\
T_i = 1 + T_{i-1} I_{\{G_{i-1} < n_t\}} \quad i = 1, 2, \ldots 
\]  

(3.2)

The chart signals an alarm when \( G_i = n_t \) and \( T_i \leq b_t \). Both counters \( G_i \) and \( T_i \) begin at 0 and they both reset when a flag is raised, that is, when \( G_i = n_t \). Note that \( n_t \) must be chosen so that it is no larger than \( b_t \), otherwise the chart would never signal. If \( n_t = b_t \), the SHDA method signals if the first \( n_t \) sets observed have sizes less than \( t_t \) or if 2\( n_t \) consecutive sets with sizes less than \( t_t \) are observed thereafter. Thus, when \( n_t = b_t \), the SHDA method closely resembles the Sets method with \( n_s = 2n_t \). Hence, when \( n_t = b_t \), the two methods become equivalent after (and if) the first flag is raised without signaling.
3.3 CUSCORE method

For the CUSCORE method, we define $X_i$, $n_c$, and $t_c$ just as we did in the Sets method, this time using the $c$ subscript to indicate the CUSCORE method. To each set size $X_i$, we assign a score $g(X_i)$ as follows:

$$g(X_i) = \begin{cases} 
1 & \text{if } X_i < t_c \\
-1 & \text{otherwise}.
\end{cases}$$

The CUSCORE (cumulative score) $C_i$ is defined as

$$C_0 = 0$$

$$C_i = \max(0, C_{i-1} I_{(C_{i-1} < n_c)} + g(X_i)) \quad i = 1, 2, \ldots.$$  \hspace{1cm} (3.3)

The CUSCORE method signals an alarm when $C_i = n_c$. The CUSCORE statistic $C_i$ defined in equation (3.3) resets after an alarm. Like the Sets method, if $n_c = 1$, the CUSCORE method reduces to a one-sided geometric Shewhart chart with a lower control limit at $t_c$. Graphically, the CUSCORE can be represented by plotting $C_i$ versus the indexes $1, 2, \ldots$ with a horizontal control limit at $n_c$. An example is shown in Figure 3.3.

![CUSCORE plot](image)

**Figure 3.3:** Example of the CUSCORE plot for $n_c = 4$ and $t_c = 1500$. Note that $C_i$ increases when $X_i < t_c$ and decreases (or stays at 0) when $X_i \geq t_c$. 
3.4 Bernoulli CUSUM

Hereafter, unless it is necessary to distinguish the Bernoulli CUSUM chart from a CUSUM chart based on some other distribution, we often refer to the Bernoulli CUSUM simply as “CUSUM.” Rather than accumulating the information into sets, the CUSUM statistic changes after each birth. Let

\[
V_j = \begin{cases} 
1 & \text{if the infant has the malformation} \\
0 & \text{if the infant does not have the malformation} \\
\end{cases} \quad j = 1, 2, \ldots .
\]

Note that \( j \) indexes each individual birth. The CUSUM statistics are given by

\[
B_0 = 0 \\
B_j = \max(0, B_{j-1}I_{\{B_{j-1} < h\}} + V_j - \delta) \quad j = 1, 2, \ldots ,
\]

where the reference value

\[
\delta = -\log \left( \frac{1-p_1}{1-p_0} \right) \\
\quad \log \left( \frac{p_1(1-p_0)}{p_0(1-p_1)} \right)
\]

arises from the Bernoulli likelihood ratio in Wald’s sequential probability ratio test. The CUSUM chart signals an alarm when \( B_j \) is greater than or equal to a control limit \( h \).

Note that the indicator function in the definition of \( B_j \) in equation (3.4) implies that the chart resets after an alarm. However, when applying the CUSUM to monitor congenital malformations, it may be more appropriate not to reset the CUSUM when it crosses the control limit \( h \) but rather continue to observe whether \( B_j \) stays above \( h \) or whether additional up crossings of the control limit occur. Such behavior would give additional evidence that the first signal was not a false alarm. This idea was suggested by Chen [7] and further developed by Kenett and Pollak [14]. For the purposes of this study, whether or not the CUSUM resets after crossing \( h \) is of lesser importance, since we use the average run length until (the first) signal as the basis to compare the four methods.

The CUSUM could be depicted graphically by plotting the CUSUM statistic \( B_j \) versus the indexes \( j = 1, 2, \ldots \) with a horizontal control limit at \( h \). However, graphing the CUSUM over extended periods of time would probably not be practical since the outcome of each of the potentially thousands of births would be plotted. A graphical example of the CUSUM chart for a relatively large value of \( p_0 \) is given in Figure 3.4.

To facilitate exact calculations of the CUSUM average run length, we adjust the size of the shift we
Figure 3.4: Example of the CUSUM chart for \( \delta = 0.15 \) and \( h = 2.7 \). Note that \( B_j \) decreases by \( \delta \) for each healthy birth and increases by \( 1 - \delta \) when a malformation is observed.

want to detect quickly \( (\gamma_1) \) by a small amount to \( \gamma_{1a} \) so that the CUSUM statistic \( B_j \) takes on a finite number of possible values. This makes it possible to calculate the average run length exactly using Markov chain theory. Specifically, we choose \( p_{1a} = \gamma_{1a}p_0 \) such that

\[
\delta = \frac{-\log \left( \frac{1-p_{1a}}{1-p_0} \right)}{\log \left( \frac{p_{1a}(1-p_0)}{p_0(1-p_{1a})} \right)} = \frac{1}{m}
\]

where \( m \) is a positive integer. The CUSUM statistics are then given by

\[
\begin{align*}
B_0 &= 0 \\
B_j &= \max \left( 0, B_{j-1}I\{B_{j-1} < h\} + V_{j-1} - \frac{1}{m} \right) \quad j = 1, 2, \ldots 
\end{align*}
\]

(3.6)

Reynolds and Stoumbos [11] described this methodology in more detail. Hence, the CUSUM is characterized by the parameters \( m \) and \( h \). An important feature of this design of the CUSUM is that \( h \) is an integer multiple of \( 1/m \). That is, \( hm = w \), where \( w \) is the number of distinct values that \( B_j \) can take without signaling.
Chapter 4

Evaluation of chart performance

4.1 Considerations in assessing chart performance

In traditional hypothesis testing, statistical tests are designed to maximize the probability of rejecting the null hypothesis when the alternative hypothesis is true, while controlling the probability that the null hypothesis is rejected incorrectly. Although control charts are not generally equivalent to hypothesis testing, there do exist some analogous concepts—especially in the monitoring phase we consider where the baseline incidence rate $p_0$ is assumed to be known. Jensen et al. [29] gave a review of control chart literature that discusses the impact of assuming that certain parameters (e.g. $p_0$) are known, when, in fact, they are estimated from historical data. Chen [21] also addressed this topic with respect to the Sets and the CUSCORE methods.

Control chart performance has traditionally been measured in terms of the average run length ($ARL$), that is, the number of data points or samples on average that are observed until the chart signals. A distinction is made between the $ARL$ under baseline and the $ARL$ under epidemic conditions. In the vernacular of the quality control literature, when $p = p_0$, the “process” is said to be “in control,” and during an epidemic when $p > p_0$, the process is said to be “out of control.” When monitoring for a change in the incidence rate, it is desirable that the control chart have a long $ARL$ when the incidence rate remains in control at the baseline level $p_0$. If the chart does signal during an in-control period, the signal is a false alarm. However, if the incidence rate increases, the control chart should signal an alarm quickly, that is, the out-of-control $ARL$ should be small.

Other types of performance criteria, such as the probability of a false alarm, the probability of successful detection, and the predictive value of an alarm were discussed by Frisén [30]. These performance metrics
were derived for the Sets method by Arnkelsdóttir [31]. However, we will use the ARL as the basis for measuring control chart performance.

4.1.1 Defining the average run length

In the context of the four methods under consideration, it is important to clarify exactly what we mean by average run length. For the Sets, CUSCORE and SHDA methods (the sets based methods), we might be inclined to define the ARL as the average number of sets that are observed until an alarm is signaled, i.e. the average number of malformed cases observed until a signal. However, in order to make valid comparisons with the Bernoulli CUSUM chart, we will define the ARL as the average number of births (or Bernoulli observations) until signal. In addition, basing the ARL on the number of births until a signal allows us to more conveniently account for the fact that a shift in the incidence rate is most likely to occur sometime between the incidences of a malformation, and not just immediately after a malformation is observed. To avoid confusion, we will use ANM to denote the average number of malformations from the onset of monitoring until signal and ANB to denote the average number of births from the onset of monitoring until signal. Sometimes we refer to the ANM or the ANB as the “initial state ANM” or the “initial state ANB” to emphasize the fact that these ARL measures assume that the chart statistics begin in the initial state (typically with values of zero). We will use the notation $ANB(p_0)$ and $ANB(p)$ to indicate the in-control (baseline) and out-of-control (epidemic) average number of births until signal, respectively. Sometimes we will simply use the term ARL if it is not necessary to make a distinction between the ANM and the ANB.

4.1.2 Steady-state average run length

The run length of any control chart depends on the starting value used for the chart statistic when monitoring begins. Naturally, if the chart starts at a value that is near the control limit, the run length will, on average, be shorter than if the chart starts at a value further from the control limit. We assume that monitoring begins during a non-epidemic period. Then, at some point in time, the shift in the incidence rate randomly occurs between birth number $\tau$ and $\tau + 1$. Note that the value of the chart statistic could be any one of its possible values when the shift occurs. A common technique used in the quality control literature to account for this situation is to assume that the chart statistic ($S_i$ for Sets method, $G_i$ and $T_i$ for the SHDA method, $C_i$ for CUSCORE, and $B_j$ for CUSUM) has reached a stationary or steady-state distribution by the time the shift occurs at $\tau$. The steady-state distribution is used to calculate a weighted average of all the possible
ARL’s that initiate from each of the possible values of the chart statistic.

The steady-state average number of births until signal \((SSANB)\) is a desirable metric because it closely resembles how the chart will perform in practice. The \(SSANB\) can be heuristically described as follows. Suppose an investigator has been monitoring for a period of time long enough to achieve the steady-state distribution under baseline conditions and that an alarm has not yet occurred. Then the shift from \(p_0\) to \(p\) occurs between births \(\tau\) and \(\tau + 1\). The average number of births until signal since the shift is the \(SSANB\).

Among all the comparisons made in the literature that is cited herein, only Wolter [8] considered steady-state ARL performance. Even so, Wolter uses the steady-state average number of malformations until a signal \((SSANM)\) to compare the Sets method with the CUSCORE method. We also note that the steady-state distribution used by Wolter is not conditional upon a false alarm not having occurred. For our purposes, using the \(SSANM\) to compare the Bernoulli CUSUM to the sets based methods is problematic because it implicitly assumes that a shift occurs immediately after a malformation, when, in fact, the shift is much more likely to occur after a normal birth because incidents are rare.

### 4.1.3 Head-start features

The Sets, CUSCORE, and SHDA methods each have a built-in head-start feature. That is, if the shift occurs prior to the onset of monitoring these methods signal faster on average than they would if the shift occurs after a period of in-control monitoring. To illustrate, suppose that after monitoring births with the Sets method for some time (assuming no alarm has yet occurred), the step-shift occurs. Let \(Z\) denote the set within which the shift occurs. Just like \(X_i\), \(Z\) can take values of \(\{0, 1, 2, \ldots\}\). (The distribution of \(Z\) is given in Section 4.5). If \(Z \geq ts\), then the counter \(S_i\) is set to 0 and one must wait until the subsequent malformation is observed in order for the chart to begin accumulating evidence that the shift has taken place. If it turns out that \(Z < ts\) (or if the shift occurs prior to the onset of monitoring), there is no delay in the accumulation of evidence for the shift.

For the SHDA method, the delay is especially problematic because in addition to the possibility that \(Z \geq ts\), if, after the shift, \(T_i > b_t\) when \(G_i\) reaches \(n_t\), the counters \(T_i\) and \(G_i\) are both reset to zero and the accumulation of evidence of the shift must begin anew. More simply stated, it is possible that at least two flags must be raised after the shift in order to signal an alarm.

The head-start feature of the SHDA method is so pronounced, in fact, that Sitter et al. [23] report that the \(ANM\) of the SHDA method is almost half that of the Sets method. The \(ANM\) used by Sitter et al. is the average number of malformations from the initial state until a signal, a performance metric that
emphasizes the advantage of the head-start of the SHDA method. The authors did not explicitly acknowledge
the head-start feature of the SHDA method. However, we show in Sections 6.2 and 6.3 that the steady-state
performance of the Sets method is, in many instances, better than that of the SHDA method.

Davis and Woodall [32] discuss the impact of using the initial state $ARL$ versus the steady-state $ARL$
when comparing methods that have head-start features. They used the steady-state $ARL$ to compare a
CUSUM and an EWMA chart (for a normal random variable) to a synthetic control chart based on runs
rules with a head start feature, similar to the SHDA method. By using the $SSARL$, they demonstrate
that once the head-start feature has worn off after a period of in-control monitoring, the CUSUM and the
EWMA outperform the synthetic control chart. The same principle applies to the sets based methods when
compared to the Bernoulli CUSUM chart.

In the following subsections, we give the derivation of the $ANB$ and the $SSANB$ for each of the
methods. All of the exact $ANB$ and $SSANB$ formulae presented below for each of the four methods (Sets,
SHDA, CUSCORE, and CUSUM) were verified by simulation.

4.2 Sets method

In this section, we discuss the derivation of the $ANM$, $ANB$, and $SSANB$ for the Sets method. In order
to calculate the $ANM$, it is necessary to first calculate $\theta_s = P(X_i < t_s)$, the probability that the set size is
less than the threshold $t_s$. Note that $\theta_s$ depends on the incidence rate $p$. Since $p_0$ is small, we can calculate
$\theta_s$ using an asymptotic approximation or we can calculate $\theta_s$ exactly using the geometric distribution.

Chen [7] demonstrated that if we express $t_s$ as a multiple of the expected value of $X_i$, that is,
$t_s = k_s \left(1 - \frac{p_0}{p_0}\right)$, and if we express the value of $p$ as a multiple of the baseline incidence rate, that is, $p = \gamma p_0$
where $0 < \gamma < 1/p_0$, then $\theta_s = P(X_i < t_s) \to 1 - e^{-\gamma k_s}$ as $p_0 \to 0$. This approximation was used by Chen
and subsequent authors who compared other methods with the Sets method [8, 9, 14, 16, 17, 18, 19]. The
approximation works well for the calculation of $ANM_s$ provided that $p_0$ is no greater than 0.01. However,
for $p_0 > 0.01$, the approximation fails rapidly. The exact value for $\theta_s$ is given by

$$\theta_s = P(X < t_s) = P(X \leq [t_s] - 1) = 1 - (1 - p)^{[t_s]} \quad (4.1)$$

where $[t_s]$ denotes the ceiling of $t_s$, that is, the smallest integer that is greater than or equal to $t_s$. Figure
4.1 shows the ratio of the exact $ANM$ calculations (using the geometric distribution to calculate $\theta_s$) to the
approximate $ANM$ (using the asymptotic approximation of $\theta_s$). The plot demonstrates that the asymptotic
approximation for the ANM loses accuracy for $p_0 > 0.01$. For each value of $p_0$ that was plotted in Figure 4.1, the Sets method was designed to be optimal at detecting a shift of $\gamma_1 = 2$ subject to $ANM_s(p_0) \geq 50$. (The optimal design of the Sets method is discussed in Section 5.2). The approximate $ANM_s(p_0)$ and the exact $ANM_s(p_0)$ were calculated for each case and their ratio plotted as the solid line ($\gamma = 1$). Likewise, the approximate $ANM_s(\gamma_1 p_0)$ and the exact $ANM_s(\gamma_1 p_0)$ were also calculated for each case and their ratio was plotted as the dashed line ($\gamma = 2$). The sharp decrease in the ratio of the $ANM_s(p_0)$ (when $\gamma = 1$) suggests that for $p_0 > 0.01$, the approximate $ANM_s$ underestimates the true value of the $ANM_s$, which would result in a higher frequency of false alarms than desired. The slight increase in the ratio of the $ANM_s(p_1)$ (when $\gamma = 2$) suggests that the approximate $ANM_s$ is overestimated, which would imply that the chart performance is worse than it actually is. Comparison of the approximate $ANM_s$ to the exact $ANM_s$ for other combinations of shift size and target ANM showed similar results. Since the comparisons of the methods presented in this work involve values of $p_0$ that are as large as 0.01, we use exact calculations given by the geometric distribution (see equation (4.1)) for all calculations of $\theta$.

We now proceed to the ANM for the Sets method. To simplify the expression in equation (4.2), we
let \( n = n_s \). Kenett and Pollak [14] showed that

\[
ANM_s(p) = \frac{1 - \theta_s^n}{\theta_s^n (1 - \theta_s)} .
\] (4.2)

Note that \( ANM_s(p) \) depends on \( p \) through \( \theta_s \) and that \( ANM_s(p) \) can be calculated for any value of \( p \). Thus, \( ANM_s(p_0) \) and \( ANM_s(p_1) \) are the in-control and out-of-control average number of malformations until signal, respectively. It is interesting to note that Page [28] derived a generalization of the \( ANM \) formula given by equation (4.2) in 1955.

To find the average number of births until signal (\( ANB \)), we must first find the expected number of births between malformations. Although the expected set size (assuming constant \( p \)) is \( E(X) = \frac{1-p}{p} \), this only includes the healthy births and omits the malformed cases. Let \( Y = X + 1 \) denote the number of consecutive normal births followed by a single malformation. Then \( Y \) is a geometric variate that takes on values \( \{1, 2, 3, \ldots\} \) with \( E(Y) = 1/p \). Using Wald’s identity, we have

\[
ANB_s(p) = ANM_s(p)E(Y) = ANM_s(p) \times \frac{1}{p} .
\] (4.3)

We initiate the discussion of how to calculate the steady-state average number of births until signal (\( SSANB \)) by focusing first on the the steady-state \( ANM \). Calculating the steady-state \( ANM \) follows directly from Markov chain theory. Relevant examples of this approach are given by Wolter [8], Reynolds and Stoumbos [11, 12], and Bourke [33, 34]. To calculate the \( ANM \) using a Markov chain, we must define the transition probability matrix \( Q_s \) whose elements \( \{q_{ij}\} \) indicate the transition probabilities of moving from state \( i \) to state \( j \). Only the transient states are represented in the transition matrix. The subscripts \( i = 1, 2, \ldots, n_s \) index the row number and \( j = 1, 2, \ldots, n_s \) index the column number of the matrix \( Q_s \). Specifically, \( q_{ij} = P(S_r = j - 1|S_{r-1} = i - 1) \) for set number \( r \). For example, if \( n_s = 4 \), the transition matrix for the Sets method is given by

\[
Q_s = \begin{bmatrix}
S_{r-1} \setminus S_r & 0 & 1 & 2 & 3 \\
0 & (1 - \theta_s) & \theta_s & 0 & 0 \\
1 & \theta_s & 0 & 0 & 0 \\
2 & (1 - \theta_s) & 0 & 0 & \theta_s \\
3 & (1 - \theta_s) & 0 & 0 & 0
\end{bmatrix}
\]

For all possible values of \( n_s > 1 \) the transition probability matrix has the same form as the one shown above.
If \( n_s = 1 \), then the Sets method reduces to a one-sided Shewhart chart with \( ANM = 1/\theta_s \). Otherwise, the \( ANM \) vector is given by \( \mathbf{N}_s = (\mathbf{I} - \mathbf{Q}_s)^{-1}\mathbf{1} \) where \( (\mathbf{I} - \mathbf{Q}_s)^{-1} \) is the fundamental matrix of the Markov chain and \( \mathbf{1} \) is a column vector of ones. Thus, in the \( ANM \) vector \( \mathbf{N}_s^T = (N_1, N_2, \ldots, N_{n_s}) \), \( N_1 \) corresponds to the average number of malformations until a signal when the chart begins in the initial state \((S_0 = 0)\), \( N_2 \) corresponds to the \( ANM \) when the chart begins in the second state \((S_0 = 1)\), and so forth.

The conditional stationary distribution, denoted by \( \mathbf{\Pi}_s^T = (\pi_1, \pi_2, \ldots, \pi_{n_s}) \), gives the probabilities that the chart is in a particular state in the distant future given that the chart has not yet signaled. In other words, \( \pi_j \) is equal to the long run proportion of time that \( S_i = j - 1 \) assuming that the chart has never signaled. More formally, let \( A_{ij} \) denote the event that \( S_i = j - 1 \) and let \( B_i \) denote that event that as of time \( i \), the chart has not signaled. That is, \( B_i = \{ \max_{1 \leq k \leq i} S_k < n_s \} \). Then

\[
\pi_j = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} I_{A_{ij} \cap B_i} = \lim_{n \to \infty} \frac{\sum_{i=1}^{n} I_{A_{ij} \cap B_i}}{\sum_{i=1}^{n} I_{B_i}}
\]

for \( j = 1, 2, \ldots, n_s \) with probability one. Note that \( \pi_j \) is invariant to the state in which the chart begins. \( \mathbf{\Pi}_s \) is given by the normalized left eigenvector corresponding to the largest eigenvalue, say \( \xi_s \), of \( \mathbf{Q}_s \). That is, \( \mathbf{\Pi}_s \) satisfies \( \mathbf{\Pi}_s^T \mathbf{Q}_s = \mathbf{\Pi}_s^T \xi_s \) and \( \mathbf{\Pi}_s^T \mathbf{1} = 1 \). The conditional stationary vector \( \mathbf{\Pi}_s \) can be found iteratively by using the power method [35].

To assess the performance of the monitoring scheme, we are interested in how long it will take for a chart to signal after the shift in the incidence rate. Since we assume that the monitoring process has been going on for some indefinite amount of time prior to the shift, we calculate the conditional stationary distribution, \( \mathbf{\Pi}_s \), under the assumption that the incidence rate is \( p_0 \). However, the \( ANM \) vector, \( \mathbf{N}_s \), is calculated assuming the incidence rate is \( p \neq p_0 \). Taking the dot product of these two vectors gives the steady state average number of malformations until a signal, \( SSANM_s = \mathbf{\Pi}_s^T \mathbf{N}_s \).

To calculate the steady-state average number of births from the shift until signal, \( (SSANB) \), we need to account for the births that occur after the shift but before the next malformation. Recall that \( Z \) represents the set in which the shift occurs. As with the sets \( X_i \), we are interested in the events \( \{ Z < t_s \} \) and \( \{ Z \geq t_s \} \). In particular, we need

\[
\beta_s = P(Z < t_s) = \frac{1}{p - p_0} \left[ p(1 - p_0)(1 - (1 - p_0)^{[t_s - 1]}) - p_0(1 - p)(1 - (1 - p)^{[t_s - 1]}) \right]
\]

Please refer to Section 4.5 for the derivation of equation (4.5). Now we define a new transition probability matrix \( \mathbf{B}_s \) that is equal to \( \mathbf{Q}_s \) except we replace \( \theta_s \) with \( \beta_s \). The steady-state average number of births until
signal is then given by

\[ SSANB_s = \frac{1}{p} (\Pi_s^T B_s N_s + 1) \]  

for any \( p \neq p_0 \). Please refer to Section 4.6 for the derivation of equation (4.6).

### 4.3 SHDA method

The initial state \( ANM \) was derived by Sitter et al.\[9\] as follows. Let \( \theta_t = P(X_i < t) \) be given by equation (4.1) except replace \( t_s \) with \( t_t \). In order to more clearly state the formulae that follow, let \( n = n_t, t = t_t, \) and \( b = b_t \). Then

\[ ANM_t(p) = 1 - \frac{\theta_t^n}{\theta_t^n (1 - \theta_t) \Psi}, \]  

where \( \Psi = \sum_{k=0}^b \psi(k) \) and \( \psi(k) \) is given by the following recursive equation:

\[
\psi(k) = \begin{cases} 
0 & \text{if } k \in \{0,1,\ldots,n-1\} \\
\theta_t^n & \text{if } k = n \\
\theta_t^n (1 - \theta_t)[1 - \sum_{j=0}^{k-n-1} \psi(j)] & \text{if } k > n.
\end{cases}
\]  

As with the Sets method, \( ANM_t(p) \) depends on \( p \) through \( \theta_t \). Applying Wald’s identity gives the \( ANB \) for the SHDA method:

\[ ANB_t(p) = ANM_t(p) \times \frac{1}{p}. \]  

The transition probability matrix for the SHDA method, \( Q_t \), is more involved than either the Sets or CUSCORE methods. For simplicity, consider an example where \( n_t = 3 \) and \( b_t = 5 \). Table 4.1 shows the Markov chain states, the corresponding values of the statistics \( G_i \) and \( T_i \) (given by equation (3.2)), and the sequences of events that give rise to those states for this example. In the third and fourth columns of Table 4.1, note that 1 represents the event \( \{X_i < t_s\} \) and 0 represents the event \( \{X_i \geq t_s\} \). The transition matrix shown in equation (4.10) corresponds to the Markov chain states that are shown in Table 4.1. The elements \( \{q_{ij}\} \) of this matrix correspond to the probability that the process moves to state \( j \) given that it is in state \( i \).
Table 4.1: States for the SHDA method with $n_t = 3$ and $b_t = 5$. In the third and fourth columns, 1 represents the event $\{X_i < t_s\}$ and 0 represents the event $\{X_i \geq t_s\}$.

<table>
<thead>
<tr>
<th>State Number</th>
<th>Value of $(G, T)$</th>
<th>Sequence(s) of possible events since last flag</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial (I)</td>
<td>$(0, 0)$ or $(3, &gt; 5)$</td>
<td>000111, 1000111, 0100111, …</td>
<td>Flag</td>
</tr>
<tr>
<td>1</td>
<td>$(0, 1)$</td>
<td>0</td>
<td>0111 or 111 to signal</td>
</tr>
<tr>
<td>2</td>
<td>$(0, 2)$</td>
<td>00, 10</td>
<td>111 to signal</td>
</tr>
<tr>
<td>3</td>
<td>$(0, \geq 3)$</td>
<td>000, 0000, 01000, 01010, …</td>
<td>111 to return to initial</td>
</tr>
<tr>
<td>4</td>
<td>$(1, 1)$</td>
<td>1</td>
<td>11 or 0111 to signal</td>
</tr>
<tr>
<td>5</td>
<td>$(1, 2)$ or $(1, 3)$</td>
<td>01, 001, 101</td>
<td>11 to signal</td>
</tr>
<tr>
<td>6</td>
<td>$(1, \geq 4)$</td>
<td>0001, 1101, 0101, 10101, …</td>
<td>11 to return to initial</td>
</tr>
<tr>
<td>7</td>
<td>$(2, 2)$, $(2, 3)$ or $(2, 4)$</td>
<td>11, 011, 0011, 1011</td>
<td>1 to signal</td>
</tr>
<tr>
<td>8</td>
<td>$(2, \geq 5)$</td>
<td>00011, 001011, 10011, …</td>
<td>1 to return to initial</td>
</tr>
<tr>
<td>Absorbing (A)</td>
<td>$(3, \leq 5)$</td>
<td>111, 0111, 00111, 10111</td>
<td>Signal</td>
</tr>
</tbody>
</table>

The Markov chain representation in Table 4.1 shows the fewest possible number of states where each row corresponds to a single state in the Markov chain. However, if we allow some of the states to be represented by more than one row in the transition matrix, it facilitates the development of a relatively simple algorithm for generating the transition matrix for any feasible combination of $n_t$ and $b_t$. Adding these extra rows does not affect the calculation $ANM$, $ANB$, $SSANM$ nor the $SSANB$. Furthermore, even with the extra rows, the Markov chain remains irreducible—a necessary condition for the existence of a unique steady-state distribution. As an example, note in Table 4.1 that state number 5 corresponds to two possible $(G, T)$ pairs: $(1,2)$ and $(1,3)$. According to the algorithm that follows, the transition probability matrix would include
separate rows for (1,2) and (1,3), even though they technically correspond to the same state in the Markov
chain.

To implement the algorithm, the states for the SHDA method can be defined by the values of the
\((G_i, T_i)\) pairs shown in Table 4.2. In Table 4.2, note that any state with \(T = b_t\) actually represents the
situation where \(T \geq b_t\). The last state \((n_t, b_t + 1)\) represents the situation where \(T \geq b_t + 1\). In addition, the
last state is also the initial state.

| Value of \((G, T)\)  |
|---|---|---|---|
| \((0, 1)\) | \((1, 1)\) | \((2, 2)\) | \((n_t - 1, n_t - 1)\) |
| \((0, 2)\) | \((1, 2)\) | \((2, 3)\) | \((n_t - 1, n_t)\) |
| \(\vdots\) | \(\vdots\) | \(\vdots\) | \(\vdots\) |
| \((0, b_t)\) | \((1, b_t)\) | \((2, b_t)\) | \((n_t - 1, b_t)\) |
| \((n_t - 1, b_t)\) | \((n_t, b_t + 1)\) |

Table 4.2: Algorithmic listing of states for the SHDA method (first read down the columns and then left to
right).

The elements of \(Q_t\) may be determined according to the following algorithm.

1. Label the rows and columns with the states’ names shown in Table 4.2. The rows correspond to the
current state, the columns correspond to the subsequent state after the next malformation is observed.

2. We generically refer to the elements of the matrix as \((g, t)(g', t')\) where the first pair indicates the row
label and the second pair indicates the column label.

3. Begin by setting all elements of \(Q_t\) equal to 0.

4. Set \((n_t, b_t + 1)(0, 1) = 1 - \theta_t\). Also set \((n_t - 1, b_t)(n_t, b_t + 1) = \theta_t\).

5. For every row \((g, t)\) where \(g < n_t\), set \((g, t)(0, \min[t + 1, b_t]) = 1 - \theta_t\). (Note that \(g\) is less than \(n_t\) for
every row except the last one).

6. If \(n_t = 1\), \(Q_t\) is now complete. If \(n_t > 1\), complete steps 7 and 8.

7. Set \((n_t, b_t + 1)(1, 1) = \theta_t\).

8. For every row \((g, t)\) where \(g < n_t - 1\), set \((g, t)(g + 1, \min[t + 1, b_t]) = \theta_t\).
The following transition matrix is an example of applying the algorithm for \( n_t = 2 \) and \( b_t = 3 \).

\[
Q_t = \begin{bmatrix}
(G_{i-1}, T_{i-1}) & (0, 1) & (0, 2) & (0, 3) & (1, 1) & (1, 2) & (1, 3) & (2, 4) \\
(0, 1) & 0 & (1 - \theta_t) & 0 & 0 & \theta_t & 0 & 0 \\
(0, 2) & 0 & 0 & (1 - \theta_t) & 0 & 0 & \theta_t & 0 \\
(0, 3) & 0 & 0 & (1 - \theta_t) & 0 & 0 & \theta_t & 0 \\
(1, 1) & 0 & (1 - \theta_t) & 0 & 0 & 0 & 0 & 0 \\
(1, 2) & 0 & 0 & (1 - \theta_t) & 0 & 0 & 0 & 0 \\
(1, 3) & 0 & 0 & (1 - \theta_t) & 0 & 0 & 0 & \theta_t \\
(2, 4) & (1 - \theta_t) & 0 & 0 & \theta_t & 0 & 0 & 0 \\
\end{bmatrix}
\]

Once \( Q_t \) is correctly defined, the steady-state average number of malformations until signal is calculated in a fashion identical to that described previously for the Sets method in Section 4.2. Recall that the conditional stationary distribution \( \Pi_t \) is calculated from \( Q_t \) assuming \( p = p_0 \) and the \( ANM \) vector \( N_t = (I - Q_t)^{-1}1 \) is calculated assuming the incidence rate is some \( p \neq p_0 \). Thus, \( SSANM_t = \Pi_t^T N_t \).

Moreover, let \( \beta_t = P(Z < t_t) \) be given by equation (4.5), replacing \( t_s \) with \( t_t \), and let \( B_t = Q_t \) except replace \( \theta_t \) with \( \beta_t \). Then

\[
SSANB_t = \frac{1}{p} (\Pi_t^T B_t N_t + 1)
\]

for any \( p \neq p_0 \). Please refer to Section 4.6 for the derivation of equation (4.11).

### 4.4 CUSCORE method

The initial state \( ANM \) for the CUSCORE method was first given by Munford [36]. As before, let \( \theta_c = P(X_i < t_c) \) with \( \theta_c \) given by equation (4.1), substituting \( t_c \) for \( t_s \). For simplicity, let \( n = n_c \). Then

\[
ANM_c(p) = \begin{cases} 
\frac{n(n+1)}{2\theta_c} & \text{if } \theta_c = \frac{1}{2}, \\
\frac{n}{2\theta_c - 1} - \frac{1 - \theta_c}{(2\theta_c - 1)^2} \left\{ 1 - \left( \frac{1 - \theta_c}{\theta_c} \right)^n \right\} & \text{otherwise.} 
\end{cases}
\]

Once again, by Wald’s identity we have

\[
ANB_c(p) = ANM_c(p) \times \frac{1}{p}.
\]
The transition probability matrix of the CUSCORE method is similar to that of the Sets method. Specifically, the elements of \( Q_c \) are given by \( q_{ij} = P(C_r = j - 1|C_{r-1} = i - 1) \) for an arbitrary time point \( r \). For example, if \( n_c = 4 \), the transition matrix is given by

\[
Q_c = \begin{bmatrix}
0 & (1 - \theta_c) & \theta_c & 0 & 0 \\
1 & (1 - \theta_c) & 0 & \theta_c & 0 \\
2 & 0 & (1 - \theta_c) & 0 & \theta_c \\
3 & 0 & 0 & (1 - \theta_c) & 0 \\
\end{bmatrix}
\]

The transition matrix will follow the same pattern as the one shown above for all \( n_c > 1 \). If \( n_c = 1 \), the CUSCORE method reduces to a one-sided Shewhart chart where \( ANM = 1/\theta_c \). Like the SHDA method, the steady-state average number of malformations until signal for the CUSCORE is calculated in a fashion identical to that of the Sets method (see Section 4.2). The conditional stationary distribution \( \Pi_c \) is calculated from \( Q_c \) assuming \( p = p_0 \) and the \( ANM \) vector \( N_c = (I - Q_c)^{-1}1 \) is calculated assuming the incidence rate is \( p \neq p_0 \). Thus, \( SSANM_c = \Pi_c^T N_c \).

Moreover, let \( \beta_c = P(Z < t_c) \) be given by equation (4.5), replacing \( t_s \) with \( t_c \), and let \( B_c = Q_c \) except replace \( \theta_c \) with \( \beta_c \). Then we have

\[
SSANB_c = \frac{1}{p} (\Pi_c^T B_c N_c + 1)
\]

for \( p \neq p_0 \). Please refer to Section 4.6 for the derivation of equation (4.14).

### 4.5 Derivation of \( \beta = P(Z < t) \) for Sets, SHDA, and CUSCORE methods

Recall that the shift occurs between \( V_\tau \) and \( V_{\tau+1} \) for \( \tau = 0, 1, 2, \ldots \). The incidence rate is \( p_0 \) for births \( j = 1, 2, \ldots, \tau \) and the incidence rate is \( p \) for \( j = \tau + 1, \tau + 2, \ldots \). Furthermore, let \( B \) index the malformation that occurs most recently before the shift. Let \( A \) index the first malformation that occurs after the shift. Hence the set \( Z \) in which the shift occurs can be represented by the sequence \( \{V_{B+1}, V_{B+2}, \ldots, V_\tau, V_{\tau+1}, \ldots, V_{A-2}, V_{A-1}\} \). Then let \( Z_0 = \tau - B \) represent the number of normal births that occur between (but not including) births \( B \) and \( \tau + 1 \). If we assume that the shift time is independent of the
It is easily verified that \( Z_0 \) can take on any value in the sequence \( \{0, 1, 2, \ldots\} \), then \( Z_0 \) is a geometric variate if we count backwards the number of births since the shift until the most recent malformation before the shift. The mass function of \( Z_0 \) is given by \( P(Z_0 = z) = p_0(1 - p_0)^z \) for \( z = 0, 1, 2, \ldots \). Let \( Z_1 = A - \tau - 1 \) represent the number of normal births after the shift that are observed before the next malformation. Then \( Z_1 \) is also a geometric random variable with mass function \( P(Z_1 = z) = p(1 - p)^z \) for \( z = 0, 1, 2, \ldots \). The set within which the shift occurs is given by \( Z = Z_0 + Z_1 \).

Note that if \( Z_0 = 0 \), the shift occurs immediately after a malformation. If \( Z_1 = 0 \), the shift occurs just prior to a malformation, and if \( Z = 0 \), the shift occurs between two successive malformed cases. By enumerating the pairs \((z_0, z_1)\) that give rise to the possible outcomes of \( Z \), we can express the mass function of \( Z \) as

\[
P(Z = z) = \sum_{j=0}^{z} p_0(1 - p_0)^j p(1 - p)^{z-j}
\]

(4.15)

for \( z = 0, 1, 2, \ldots \), since we assume \( Z_0 \) and \( Z_1 \) are independent. Assuming \( p_0 \neq p \), we have \( \phi \equiv \frac{1-p_0}{1-p} \neq 1 \). With a little algebra and the well known identity for the sum of a finite geometric series, we have

\[
P(Z = z) = \sum_{j=0}^{z} p_0(1 - p_0)^j p(1 - p)^{z-j} \\
= p_0 p(1 - p)^z \sum_{x=0}^{z} \phi^x \\
= p_0 p(1 - p)^z \left[ \frac{1 - \phi^{z+1}}{1 - \phi} \right] \\
= \frac{p_0 p}{p - p_0} [(1 - p_0)^{z+1} - (1 - p)^{z+1}] \\
= \frac{p_0 p}{p - p_0} [(1 - p_0)^z - (1 - p)^z] .
\]

(4.16)

It is easily verified that \( \sum_{z=0}^{\infty} \frac{p_0 p}{p - p_0} [(1 - p_0)^z - (1 - p)^z] = 1 \). We now use equation (4.16) to calculate \( \beta = P(Z \leq t) \). Recall that \( \lceil t \rceil \) denotes the ceiling of \( t \), that is, the smallest integer that is greater than or equal to \( t \).

\[
\beta = P(Z \leq \lceil t \rceil - 1)
= \sum_{z=0}^{\lceil t \rceil - 1} \frac{p_0 p}{p - p_0} [(1 - p_0)^z - (1 - p)^z] \\
= \frac{p_0 p}{p - p_0} \left[ (1 - p_0) \left( \frac{1 - (1 - p_0)^{\lceil t \rceil}}{p_0} \right) - (1 - p) \left( \frac{1 - (1 - p)^{\lceil t \rceil}}{p} \right) \right] \\
= \frac{1}{p - p_0} \left[ p(1 - p_0)(1 - (1 - p_0)^{\lceil t \rceil}) - p_0(1 - p)(1 - (1 - p)^{\lceil t \rceil}) \right] .
\]

(4.17)
The assumption of $p_0 \neq p$ that is required for equations (4.16) and (4.17) does not pose a problem in our study, since $\beta$ is only used to calculate the SSANB for $p > p_0$.

4.6 Derivation of SSANB for Sets, SHDA, and CUSCORE Methods

We give below a general derivation of the steady-state average number of births since the shift until signal (SSANB) which is applicable for the Sets, SHDA, and CUSCORE methods. To calculate the SSANB we must remember that it is very unlikely that the shift occurs immediately following a malformation since $P(V_\tau = 1) = p_0$ and $p_0$ is small. This necessitates accounting for the births $\{V_{\tau+1}, V_{\tau+2}, \ldots, V_{A-1}, V_A\}$ that occur between the shift and the next observed malformation. (Recall that $A$ indexes the first malformation to occur after the shift). Let $Y_1 = A - \tau$ be the number of normal births observed after the shift until (and including) the next malformation. Note that $Y_1$ is a geometric variate that takes on values $\{1, 2, \ldots\}$ and has expectation $1/p$. Let $R$ be the number of births that occur after birth $A$ until the chart signals. Thus, the total number of births from the shift until signal is $Y_1 + R$.

To develop the ideas heuristically, suppose we have been monitoring for some time using the Sets method with $n_s = 5$ and that, prior to any signal, a shift occurs sometime between malformation $z$ and malformation $z + 1$. Suppose also that the value of the counter just prior to the shift is $S_z = 3$. The value of $S_{z+1}$ depends on whether the set in which the shift takes place, $Z$, has a value less than $t_s$ or not. If $Z < t_s$, then $S_{z+1} = 4$ and the expected number of births from shift until signal is $E(Y_1) + E(R|S_{z+1} = 4) = \frac{1}{p} + \frac{N_5}{p}$. (Recall $N_i = ANM$ given that $S_0 = i-1$). Because we assume the steady-state distribution has been achieved and since the value of $Z$ does not depend on the previously observed $\{X_1, X_2, \ldots, X_z\}$, the event $\{S_z = 3$ and $Z < t_s\}$ occurs with probability $\pi_4\beta_s$ ($\pi_4$ is defined by equation (4.4)). Similarly, we can consider the event $\{S_z = 3$ and $Z \geq t_s\}$ which occurs with probability $\pi_4(1 - \beta_s)$ and has corresponding expected number of births from shift until signal equal to $E(Y_1) + E(R|S_{z+1} = 0) = \frac{1}{p} + \frac{N_1}{p}$. The SSANB is given by the weighted average of the expected number of births from shift until signal resulting from each of the possible combinations of the state prior to the shift and the dichotomous outcomes $\{Z < t_s\}$ and $\{Z \geq t_s\}$, with the corresponding probabilities $\pi_k\beta_s$ and $\pi_k(1 - \beta_s)$ for $k = 1, \ldots, n_s$ as the weights.

To generalize the approach for all the sets based methods, let $k = 1, 2, \ldots, w$ index the Markov chain states where $w$ is the number of transient states. Let $N^-_k$ denote the expected number of malformations after the event $\{Z < t\}$ is observed until signal when the state prior to the shift was $k$. Similarly, let $N^+_k$ denote
the expected number of malformations after the event \( \{ Z \geq t \} \) is observed until signal when the state prior to the shift was \( k \). The probability corresponding to \( N_k^+ \) is \( \pi_k \beta \) and the probability corresponding to \( N_k^- \) is \( \pi_k(1 - \beta) \), where \( \pi_k \) is the long run proportion of times that the chart is in state \( k \) given that the chart has not yet signaled. Also note that for at least one of the states, the observation of \( \{ Z < t \} \) will result in an immediate signal—and thus \( N_k^- = 0 \) for these states. In general, for any given state \( K \) prior to the shift, the \( ANB \) after the shift is

\[
\text{ANB} = \mathbb{E}(Y_1) + \mathbb{E}(R|K, Z) = \frac{1}{p} + \frac{N_k^+}{p} \quad \text{or} \quad \frac{1}{p} + \frac{N_k^-}{p},
\]

depending on whether \( \{ Z < t \} \) or \( \{ Z \geq t \} \), respectively. Taking the sum of all the possible values of \( ANB \) multiplied by their corresponding probabilities gives

\[
SSANB = \sum_{k=1}^{w} \pi_k (1 - \beta) \left( \frac{1}{p} + \frac{N_k^+}{p} \right) + \sum_{k=1}^{w} \pi_k \beta \left( \frac{1}{p} + \frac{N_k^-}{p} \right)
= \frac{1}{p} \sum_{k=1}^{w} \pi_k \left( (1 - \beta)(1 + N_k^+) + \beta(1 + N_k^-) \right)
= \frac{1}{p} \left( \sum_{k=1}^{w} \pi_k \left( (1 - \beta)N_k^+ + \beta N_k^- \right) + 1 \right),
\]

where, with a bit of algebra, the last equality in equation (4.18) results from the fact that \( \sum_{k=1}^{w} \pi_k = 1 \).

Recall that \( B = Q \) except that \( \theta \) is replaced by \( \beta \). Note that \( B \) is the transition probability matrix for the set \( Z \) in which the shift occurs. Let \( B^{(k)} \) denote the \( k \)th row of \( B \) for \( k = 1, 2, \ldots, w \). Also recall that \( N \) is the vector of \( ANM \) values that correspond to each of the Markov chain states. Then for the Sets method, the SHDA method, and the CUSCORE method, examination of the matrix \( B \) readily shows that \( B^{(k)}N = (1 - \beta)N_k^+ + \beta N_k^- \) for every \( k \). Thus, taking the inner portion of the last line in equation (4.18), we have

\[
\sum_{k=1}^{w} \pi_k \left( (1 - \beta)N_k^+ + \beta N_k^- \right) = \sum_{k=1}^{w} \pi_k B^{(k)}N = \Pi^T BN,
\]

and substitution of equation (4.19) into equation (4.18) gives

\[
SSANB = \frac{1}{p} \left( \Pi^T BN + 1 \right).
\]

### 4.7 Bernoulli CUSUM

Because the CUSUM does not group the individual outcomes into sets, we only consider the \( ANB \) and \( SSANB \) for the CUSUM chart. These quantities are calculated using the Markov chain methodology. In the Appendix of their paper, Reynolds and Stoumbos [11] provided a description of the states of the Markov
chain and how to construct the transition probability matrix. In what follows, we reproduce their description with minor changes, adapting the notation and terminology to match this work.

According to the definition of the CUSUM statistic given in equation (3.6), $B_j$ can take on the following values without signaling: $\{\frac{k-1}{m} : k = 1, 2, \ldots, w\}$. Let $k$ index the states of the Markov chain such that when we refer to $k$, we refer to the state when $B_j = \frac{k-1}{m}$. Let $w = mh$ be the number of transient states in the Markov chain. Since $V_j$ is a Bernoulli variate, the possible values for $(V_j - 1/m)$ are $-1/m$ and $(m-1)/m$. Thus, after each birth the transition will be either down 1 state (for $V_j = 0$) or up $(m-1)$ states (for $V_j = 1$). Recall that since $m$ is an integer, $h$ can be taken to be an integer multiple of $1/m$. Thus, if $h < 1$ (which implies $h \leq (m-1)/m$), then from any state a signal will be given if a malformation is observed. This means that the CUSUM chart reduces in this case to a rule which signals as soon as a malformation is observed, so the ANB is simply $1/p$ where $p = P$(observing a malformation). Thus, in what follows, we assume that $h \geq 1$. Let $c = w - m + 1$ represent the highest state from which it is not possible to signal. This means that if the current state is $k \leq c$, then observing a malformation will result in a transition to state $k + m - 1 \leq w$, but if the current state is $k > c$, then observing a malformation will result in a signal. If $h \geq 1$, then this implies that $c \geq 1$. The transition probability matrix, $Q_b$, for this Markov chain is relatively simple. In particular, if $q_{ij}$ is the transition probability from state $i$ to state $j$ (and the $(i,j)$th element of $Q_b$) then

\[
q_{11} = 1 - p, \\
q_{1m} = p, \\
q_{i,i-1} = 1 - p \quad \text{for } i = 2, 3, \ldots, w, \\
q_{i,i+m-1} = p \quad \text{for } i = 2, 3, \ldots, c, \text{ if } c \geq 2, \\
q_{ij} = 0 \quad \text{otherwise.}
\]

(4.21)
Below is an example of the transition probability matrix for the CUSUM with $m = 4$ and $h = 2$.

\[
\begin{pmatrix}
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\
1 & (1 - p) & 0 & 0 & p & 0 & 0 & 0 \\
2 & (1 - p) & 0 & 0 & 0 & p & 0 & 0 \\
3 & 0 & (1 - p) & 0 & 0 & 0 & p & 0 \\
4 & 0 & 0 & (1 - p) & 0 & 0 & 0 & p \\
5 & 0 & 0 & 0 & (1 - p) & 0 & 0 & 0 \\
6 & 0 & 0 & 0 & 0 & (1 - p) & 0 & 0 \\
7 & 0 & 0 & 0 & 0 & 0 & (1 - p) & 0 \\
8 & 0 & 0 & 0 & 0 & 0 & 0 & (1 - p)
\end{pmatrix}
\]

The ANB vector is given by $N_b = (I - Q_b)^{-1}1$ where $N_b^T = (N_1, N_2, \ldots, N_w)$. Thus, the initial state average number of births until signal is given by

\[
ANB_b(p) = N_1 .
\]  
(4.22)

As with the previous methods, the steady-state distribution vector $\Pi_b$ is the normalized left eigenvector corresponding to the largest eigenvalue of $Q_b$, calculated assuming $p = p_0$. $N_b$ is calculated assuming $p = \gamma p_0$, and the steady-state average number of births until signal is given simply by

\[
SSANB_b = \Pi_b^T N_b .
\]  
(4.23)

Unlike the sets based methods, the CUSUM typically has hundreds and potentially thousands of states. This is because as $p_0$ gets smaller, the number of states required for the Markov chain becomes larger. Thus, as the dimension of $Q_b$ increases, it becomes increasingly difficult to invert $I - Q_b$. One way to circumvent this problem is to calculate $N_b$ using explicit algebraic solutions and thus avoid any matrix inversion.

Reynolds and Stoumbos [11] published explicit algebraic solutions for the ANB vector $N_b$. These proved very useful for calculating $ANB_b(p_0)$, a necessary step for designing the CUSUM charts (see Chapter 5). However, we discovered that the explicit solutions, while algebraically correct, can be numerically unstable. For calculations of $ANB_b(p_0)$, when $\gamma = p/p_0 = 1$, the explicit solutions provide correct results almost instantaneously. However, as $\gamma$ increases, the results computed from the explicit solutions become affected by the numerical instability. For this reason, we only used them to calculate the in-control $ANB_b(p_0)$.
In addition, the explicit solutions given by Reynolds and Stoumbos [11] contain some minor typographical errors. In Section 4.8 we present a corrected revision of the explicit solutions, which, unfortunately, are still prone to numerical instability for larger values of \( \gamma \). The alternative to computing the SSANB is the use of simulation. An algorithm for the simulation of SSANB of the CUSUM is given below in Section 4.9.

4.8 Revised explicit solutions of the ANB for the CUSUM

We reproduce below the work of Reynolds and Stoumbos [11], adapting the terminology and notation to match this work, along with some minor corrections. In what follows below, let \( p \) be the incidence rate and \( q = 1 - p \).

The transition probability matrix \( Q_b \) determined by (4.21) has a simple structure, so it is possible to solve directly for the ANB, although the resulting expression is rather complicated. The direct solution will be given here for the case of \( p_0 \leq p_1 \). Although these expressions are algebraically correct and known to be reliable when \( p = p_0 \), they are subject to numerical instability as \( \gamma = p/p_0 \) increases. Using the fact that \( N_b = (I - Q_b)^{-1}1 \), it is a simple matter to show that \( N_b \) satisfies the equation

\[
N_b = 1 + Q_b N_b .
\]  

(4.24)

Recall \( c = w - m + 1 \) and \( N_b^T = (N_1, N_2, \ldots, N_w) \). Consider the first case where \( c \geq m \). Writing out the system of equations in (4.24) gives

\[
\begin{align*}
N_1 & = 1 + q N_1 + p N_m \\
N_2 & = 1 + q N_1 + p N_{m+1} \\
N_3 & = 1 + q N_2 + p N_{m+2} \\
& \vdots \\
N_c & = 1 + q N_{c-1} + p N_w \\
N_{c+1} & = 1 + q N_c \\
N_{c+2} & = 1 + q N_{c+1} \\
& \vdots \\
N_w & = 1 + q N_{w-1} .
\end{align*}
\]  

(4.25)
The approach used to solve these equations is to find $N_c$; then for $i \neq c$, obtain $N_i$ from $N_c$. For $i > c$, it is relatively easy to show using the equations in (4.25) that

$$N_i = \frac{1 - q^{i-c}}{p} + q^{i-c}N_c. \quad (4.26)$$

Once $N_c$ and $N_i$ for $i > c$ are obtained, $N_i$ for $i < c$ can be obtained by finding $N_{c-1}, N_{c-2}, \ldots, N_2$ successively, using the equation

$$N_i = \frac{1}{q} (N_{i+1} - pN_{i+m} - 1). \quad (4.27)$$

Equation (4.27) is obtained from the equation $N_{i+1} = 1 + qN_i + pN_{i+m}$, which holds for $i = 2, 3, \ldots, c$. An expression for $N_1$ obtained from the first equation in the system of equations in (4.25), namely

$$N_1 = \frac{1}{p} (1 + pN_m). \quad (4.28)$$

The expression for $N_c$ is obtained by working up through the system of equations in (4.25), with each $N_i$ being expressed in terms of $N_{i+1}, N_{i+2}, \ldots$. Once $N_c$ is reached in the sequence, equation (4.26) can be used to give an equation with terms involving only $N_c$. This equation can then be solved explicitly for $N_c$. The resulting solution can be expressed using a set of constants $T(i', i)$ defined recursively by

$$T(0, i) = \begin{cases} 
0 & \text{if } i = -1, 0 \\
1 & \text{if } i = 1, 2, \ldots, m-1,
\end{cases}$$

$$T(1, i) = \begin{cases} 
b & \text{if } i = -1 \\
b - ia & \text{if } i = 0, 1, \ldots, m-1,
\end{cases}$$

and for $i' \geq 2$,

$$T(i', i) = \begin{cases} 
T(i' - 1, m-1) & \text{if } i = -1 \\
T(i' - 1, m-1) - a \sum_{j=0}^{i} T(i' - 1, j) & \text{if } i = 0, 1, \ldots, m-1.
\end{cases} \quad (4.29)$$

where $a = pq^{m-1}$ and $b = 1 - q^{m-1}$. Now, fix $i' \in \{1, 2, \ldots\}$ and $k \in \{0, 1, \ldots, m-1\}$ such that $c = i'm + k$. Then we have

$$N_c = \frac{\sum_{j=1}^{4} \kappa_j}{T(i'+1, k-1)}. \quad (4.30)$$
where

\[ \kappa_1 = \begin{cases} 0 & \text{if } k = 0 \\ 2 \sum_{i=0}^{k-1} q^i T(i', k - 1 - i) & \text{otherwise,} \end{cases} \]

\[ \kappa_2 = \begin{cases} 0 & \text{if } i' \leq 1 \\ q^k \sum_{i=0}^{m-1} q^i \left[ \sum_{j=0}^{i'-2} q^{jm} T(i' - 1 - j, m - 1 - i) \right] & \text{otherwise,} \end{cases} \]

\[ \kappa_3 = \begin{cases} 0 & \text{if } k = m - 1 \text{ or } i' = 0 \\ q^k \sum_{i=0}^{m-1-k-1} q^i T(i' - 1, m - 1 - i) & \text{otherwise,} \end{cases} \]

\[ \kappa_4 = \frac{1}{p} \left( q^{(i'-1)m+k} - T(i', k) + T(i' + 1, k - 1) \right). \]

For the case in which \(1 \leq c < m\), a derivation similar to the one above gives

\[ N_c = \frac{1 + T(1, c - 1)}{p T(1, c - 1)} \]

and the other values of \(N_i\) are calculated using equations (4.25), (4.26), (4.27), and (4.28) as described above. And lastly, for the trivial case in which \(h < 1\), we have \(N_1 = N_2 = \ldots = N_w = 1/p\).

We finish with a practical note regarding the calculation of the coefficients \(T(i', i)\). The coefficients can be calculated recursively by storing their values progressively in a matrix with \((i' + 2)\) rows and \((m + 1)\) columns. Here the rows of the matrix correspond to the first index, \(i'\) and the columns to the second index, \(i\).
4.9 Algorithm for the simulation of the CUSUM SSANB

To vastly improve the speed of the simulation, geometric random variables were simulated instead of Bernoulli random variables. This is possible because 1) there is no loss of information in only observing the number of births between malformations, and 2) the CUSUM can only signal after the observation of a malformation.

In addition, the Markov chain states were simulated (using integers as labels) as opposed to the control statistic $B_j$. Since the CUSUM signals when $B_j = h$, we found that if we simulated the control statistic, then $B_j$ would occasionally not be exactly equal to $h$ (when it should have been) due to machine rounding error, giving incorrect results.

The strategy is to simulate the in-control process for an extended period of time (up to $1.5 \times ANB_0$ births), where we are confident that the control statistic has reached the steady-state distribution. Our studies suggest that convergence to the steady-state distribution occurs long before $1.5 \times ANB_0$ (see Section 7.1.2). Then we introduce the increase of the incidence rate from $p_0$ to $p$ and count the number of births since the shift until signal. This is repeated independently and the average of the run lengths and the standard error of the average are calculated and used to estimate the SSANB. If a simulated sequence signals prior to reaching the shift time, it is discarded and a new replication begins.

Before giving the algorithm, let us define the following variables:

\[ \tau = \lceil 1.5 \times ANB_0 \rceil \]
\[ m \text{ is a positive integer which equals } (1/\text{reference value}) \text{ of the CUSUM} \]
\[ w = mh \text{ is the total number of transient Markov chain states} \]
\[ n = \text{run length counter (number of births)} \]
\[ k = \text{indicates the state after the most recently observed malformation, taking values } \{1, 2, \ldots, w\} \]
\[ k_p = \text{indicates the state after the second most recently observed malformation} \]

The steps of the simulation are as follows:

1. Calculate $\tau$.

2. Initialize variables: set $n=0$ and $k=1$.

3. Set $k_p = k$.

4. Generate $x$, a single realization of the geometric random variable $X$ under incidence rate $p_0$. $X$ represents the number of normal births between (but not including) malformations, and can take
on values \( \{0, 1, 2, \ldots\} \).

5. Let \( k = \max(1, k - x) + m - 1 \). Increment the run-length counter to \( n = n + x + 1 \).

6. If at any time, \( (k > w) \) and \( (n \leq \tau) \), a false alarm has occurred, in which case return to step (2).

7. If \( n \leq \tau \), return to step (3) and repeat. Otherwise go to step (8).

8. Since \( n > \tau \), we have actually passed the shift time. Now we backtrack to the state following the second previous malformation by subtracting off the births that moved the run-length past the shift time. Therefore, let \( n = n - x - 1 \).

9. Now we advance the previous state counter by accounting for the \( \tau - n \) births that occurred between the second to last malformation and the shift. We do this by letting \( k = \max[1, k_p - (\tau - n)] \).

10. Reset the run-length counter by letting \( n = 0 \).

11. Generate \( x \), a single realization of the geometric random variable \( X \) under incidence rate \( p \). Note \( X \) can take on values \( \{0, 1, 2, \ldots\} \).

12. Let \( k = \max(1, k - x) + m - 1 \). Increment the run-length counter to \( n = n + x + 1 \).

13. If \( k \leq w \), return to step (11) and repeat. Otherwise, go to step (14).

14. Now that \( k > w \), a signal occurs and the replicate is complete. Record the run-length \( n \) and begin a new replication by returning to step (2).

4.10 Weighted average of the \textit{ARL}'s for an arbitrary shift time

The primary assumption when using the \textit{SSANB} as a performance metric is that the steady-state distribution has been achieved prior to occurrence of the shift. To assess the impact of this assumption, we begin by deriving \( \text{WANB}(\tau) \), the weighted average of the average number of births from an arbitrary shift time \( \tau \) until signal. \( \text{WANB}(\tau) \) is calculated in a fashion similar to the \textit{SSANB}, except that the distribution of the Markov chain states that arises when the shift occurs at \( \tau \) is used instead of the conditional steady-state distribution. Naturally, \( \lim_{\tau \to \infty} \text{WANB}(\tau) = \text{SSANB} \).

For the sets based methods, the \textit{SSANB} is given by equation (4.20). In this equation, the conditional stationary distribution, \( \Pi \), can be interpreted as the probability that the Markov chain is in one of the corresponding states after the chart has been running long enough to achieve the steady-state distribution,
assuming that no alarm has taken place. We can, however, examine the probability distribution of being in
the various states after \( i = -1, 0, 1, 2, \ldots \) malformations or after \( j = 0, 1, 2, \ldots \) births given that we began
in the initial state of the chart and that no alarm has yet occurred. For the three sets based methods, let
the value \( i = -1 \) indicate that the shift takes place before monitoring begins. Let \( i = 0 \) indicate that the
shift takes place between the start of monitoring and the observance of the first malformation and let \( i \geq 1 \)
indicate that the shift has taken place between malformations \( i \) and \( i + 1 \). Let \( \Pi_i^{(i)}, l = \{ s, c \} \), denote the
“\( i \)-step” distribution (the distribution of the Markov chain states after malformation \( i = 0, 1, \ldots \)) of the Sets
and CUSCORE methods, respectively. Then

\[
(\Pi_i^{(i)})^T = \begin{cases} [1, 0, 0, \ldots] & \text{if } i = 0, \\ Q_i^l & \text{if } i = 1, 2, \ldots \end{cases}
\] (4.31)

for \( l = \{ s, c \} \). In equation (4.31), \( Q_i^l \) represents the matrix multiplication of the transition probability
matrix \( Q_l \) by itself \( i \) times and \( [M]_{\text{first}} \) denotes the first row of the matrix \( M \) whose elements have been
standardized so that they sum to unity. Since the initial state of the SHDA method is given by the last row
of \( Q_t \), the \( i \)-step distribution is found by standardizing the last row of \( Q_i^l \). Thus

\[
(\Pi_i^{(i)})^T = \begin{cases} [\ldots, 0, 0, 1] & \text{if } i = 0, \\ [Q_i^l]_{\text{last}} & \text{if } i = 1, 2, \ldots \end{cases}
\] (4.32)

where \( [M]_{\text{last}} \) denotes the last row of the matrix \( M \) whose elements have been standardized so that they
sum to unity. In equations (4.31) and (4.32), when \( i = 0 \), the number of zeros included in the vector is one
less than the number of Markov chain states. For the sets based methods, it follows that for a shift of size \( \gamma \),

\[
WANB_i^{(i)}(\gamma) = \begin{cases} ANB_i(\gamma p_0) & \text{if } i = -1, \\ \frac{1}{\gamma p_0} (\Pi_i^{(i)})^T B_i N_t + 1 & \text{if } i = 0, 1, \ldots \end{cases}
\] (4.33)

for \( l = \{ s, t, c \} \). \( WANB_i^{(i)}(\gamma) \) is thus the weighted average of the average number of births since a shift that
takes place between incidents \( i \) and \( i + 1 \) until signal.

For the CUSUM, let \( j = 0 \) indicate that the shift takes place before the onset of monitoring and \( j \geq 1 \)
indicate that the shift takes place between births \( j \) and \( j + 1 \). Let \( \Pi_j^{(j)} \) denote the “\( j \)-step” distribution of
the CUSUM (the distribution of the Markov chain states after birth \( j = 0, 1, \ldots \)). Then

\[
(\Pi_b^{(j)})^T = \begin{cases} 
[1, 0, 0, \ldots] & \text{if } i = 0, \\
Q_0^{(j)}_b & \text{if } i = 1, 2, \ldots
\end{cases} \tag{4.34}
\]

and

\[
WANB_b^{(j)}(\gamma) = (\Pi_b^{(j)})^T N_b \quad \text{for } j = 0, 1, 2, \ldots \tag{4.35}
\]

\(WANB_b^{(j)}\) is thus the weighted average of the average number of births since a shift that takes place between births \( j \) and \( j + 1 \) until signal. Implicit in equation (4.35) is that \( WANB_b^{(0)}(\gamma) = ANB_b(\gamma p_0)\). In equations (4.31), (4.32), and (4.34), the transition probability matrix \( Q \) is constructed assuming \( p = p_0 \). In equation (4.33), \( B_l \) is equal to \( Q_l \) with \( \theta_l \) replaced by \( \beta_l \) for \( l = \{s, t, c\} \). In equations (4.33) and (4.35), \( N_l \) is calculated assuming \( p = \gamma p_0 \) for \( l = \{s, t, c, b\} \).

### 4.11 Measuring \(SSANB\) performance across a range of possible shifts

Since the magnitude of the observed shift (\( \gamma \)) is not likely to be the shift which the chart was optimally designed to detect (\( \gamma_1 \)), it is desirable that a chart have good performance at values of \( \gamma \) that differ from \( \gamma_1 \). If we are willing to assume that the possible shifts \( \gamma \) follow a known distribution \( \Gamma \), we could measure the chart’s performance using

\[
\int SSANB(\gamma) d\Gamma . \tag{4.36}
\]

If we assume that the possible shifts are uniformly distributed on the interval \([a, b]\), then the integral in (4.36) is proportional to

\[
\int_a^b SSANB(\gamma) d\gamma . \tag{4.37}
\]

In practice, we can only approximate the integral in (4.37) by evaluating \(SSANB(\gamma)\) at chosen points in the interval \([a, b]\) using quadrature or some other integral approximation. The function \(SSANB(\gamma)\) decreases very rapidly as \( \gamma \) increases from unity. However, \(\log(SSANB(\gamma))\) decreases less rapidly than \(SSANB(\gamma)\).
and, as a result, the “curve” formed by joining the evaluation points with straight lines is smoother when we take the natural logarithm of the SSANB. Hence, for the same chosen values of $\gamma$ in $[a, b]$, the approximation of $\int_a^b \log(SSANB(\gamma))d\gamma$ will be more accurate than the approximation of $\int_a^b SSANB(\gamma)d\gamma$. In Section 6.3 we use $\int_a^b \log(SSANB(\gamma))d\gamma$ to compare the performance of the four methods.
Chapter 5

Determining the design parameters

5.1 Optimality criteria

Prior to implementation of any surveillance scheme, the design parameters are determined with the goal of avoiding false alarms while also maximizing the responsiveness of the chart to a shift. For example, with the Sets method, we would like to choose specific values for \( n_s \) and \( t_s \) that will optimize the chart’s responsiveness to a shift. In the comparisons that follow in Chapter 6, design parameters are chosen to be optimal in the sense used by Lorden [37] and Moustakides [38]. Simply stated, this criteria selects the values of the design parameters that minimize the worst case \( ANB(p_1) \) subject to the constraint that \( ANB(p_0) \) is at least as large as a target in-control average number of births until signal, \( b_0 \). For convenience in choosing the value of \( b_0 \), we also define \( m_0 = p_0 b_0 \), the target in-control average number of malformations until signal. The worst case \( ANB(p_1) \) arises when the shift takes place at a moment that least favors the detection of the shift.

More formally, let the random variable \( N \) represent the run length of the chart. For \( \tau = 1, 2, \ldots \), define the independent, sequentially observed Bernoulli random variables \( V_1, V_2, \ldots, V_{\tau-1} \) with incidence rate \( p_0 \) and \( V_\tau, V_{\tau+1}, \ldots \) with incidence rate \( p_1 \). Let \( P_\tau \) be the joint distribution of \( V_1, V_2, \ldots, \) and \( E_\tau \) the expectation under \( P_\tau \). Then define the worst case average number of births since shift until signal as

\[
W = \sup_{\tau \geq 1} \text{ess sup} E_\tau[(N - \tau + 1)^+|V_1, \ldots, V_{\tau-1}] .
\] (5.1)

Informally, the essential supremum of the conditional expectation in equation (5.1) can be interpreted as the largest possible \( ANB \) that arises from the chart being in the least advantageous state when the shift
occurs at time $\tau$. The $\sup_{\tau \geq 1}$ identifies the largest $ANB$ resulting from the shift occurring at the least advantageous time $\tau$. The objective is to choose the design parameters of the chart that minimize $W$ subject to $ANB(p_0) \geq b_0$.

Moustakides [38] proved that Page’s [10] stopping rule (the CUSUM based on Wald’s sequential probability ratio test) is optimal in this sense. In the context of our study, this means that among the class of surveillance schemes that achieve $ANB(p_0) \geq b_0$, the Bernoulli CUSUM has the lowest possible $W$ for the shift $\gamma_1$ that the chart was designed to detect. In order to make comparisons with the CUSUM as equitable as possible, we selected the design parameters of the sets based methods to also minimize $W$ subject to $ANB(p_0) \geq b_0$.

### 5.2 Sets method

Recall that to implement the Sets method, the design parameters $n_s$ and $t_s$ must be specified. To assist in determining the worst possible $ANB$ for the Sets method, we assume that the shift takes place sometime between the malformation $z$ and malformation $z + 1$. For the Sets method, the worst possible $ANB$ from shift until signal occurs when the shift happens after the currently accumulating set has already crossed the threshold $t_s$. Restated symbolically, when the sequence of births prior the shift, $V_1, V_2, \ldots, V_{\tau - 1}$, results in $Z_0 \geq t_s$, the worst possible $ANB$ will occur ($Z_0$ is defined in Section 4.5). This is because $Z_0 \geq t_s$ implies that $S_{z+1} = 0$. Consequently, the first malformation observed after the shift does not contribute to the accumulating evidence that the incidence rate has increased. As in Section 4.6, let $Y_1$ be the number of births after the shift until (and including) the next malformation. Define $R$ as the number of births that occur after the first malformation following the shift until the chart signals. The worst case $ANB$ for the Sets method is given by

$$W_s = E(Y_1) + E(R \mid S_{z+1} = 0) = \frac{1}{p_1}[1 + ANM_s(p_1)]$$

(5.2)

where $ANM_s(p_1)$ is given by equation (4.2). Since the only portion of $W_s$ that depends on the design parameters $n_s$ and $t_s$ is $ANM_s(p_1)$, the values of $n_s$ and $t_s$ that minimize $W_s$ are equivalent to those that minimize $ANM_s(p_1)$.

Incidentally, Chen proposed various approaches for choosing the parameters for the Sets method [7, 25, 39]. These approaches were challenged by Gallus et al. [17] and Sitter et al. [9] in favor of choosing the parameters that minimize $ANM_s(p_1)$ subject to $ANM_s(p_0) \geq m_0$ for a specified value of $\gamma_1$. Gallus et
al. [17] set forth an iterative algorithm for identifying the optimal parameter choices for the Sets method, while Sitter et al. [9] advocated a grid search. The iterative algorithm and the grid search both give rise to the same values for \( n_s \) and \( b_s \) given the choice of \( p_0, \gamma_1, \) and \( m_0 \). For our study we used a grid search based on the following steps to choose the Sets method parameters \( n_s \) and \( t_s \).

1. Begin with \( n_s = 0 \).

2. Let \( n_s = n_s + 1 \). Note that for fixed \( n_s \) and \( p, \ ANM_s(p) \) is a decreasing function of \( t_s \). Using equation (4.2), calculate \( ANM_s(p_0) \) for various integer values of \( t_s \) until a value of \( t_s \) is found such that \( ANM_s(p_0) \geq m_0 \) but \( t_s + 1 \) results in \( ANM_s(p_0) < m_0 \).

3. Calculate and record \( ANM_s(p_1) \) using the values of \( n_s \) and \( t_s \) found in step (2).

4. Return to step (2) and repeat until there is reasonable assurance that increasing values of \( n_s \) only results in increasing values \( ANM_s(p_1) \).

5. Choose the \((n_s, t_s)\) combination that results in the smallest \( ANM_s(p_1) \).

Gallus et al. [17] noted that simulation results suggest that the \( ANM_s(p_1) \) has a “unique minimum” with respect to \( n_s \). We also found this to be true, even though the \( ANM_s(p_1) \) as a function of \( n_s \) is not always smooth—owing to the discreteness of the geometric variables and the corresponding values of \( t_s \) needed to achieve \( ANM_s(p_0) \geq m_0 \) for each value of \( n_s \). For this reason we suggest in step (4) above that the values of \( n_s \) should be incremented until the increasing pattern of \( ANM_s(p_1) \) is clear and the minimum obviously determined. Figure 5.1 demonstrates graphically how the optimal choices for \( n_s \) and \( t_s \) are determined. Note that in Figure 5.1(B), each value of \( ANM_s(0.002) \) is calculated using the corresponding value of \( t_s \) that is shown in Figure 5.1(A). The convex shape of the \( ANM_s(p_1) \) curve shown in Figure 5.1(B) was typical for the various cases of the Sets method that we examined.

### 5.3 SHDA method

For the SHDA method, the parameters \( n_t, b_t, \) and \( t_t \) must be specified. To arrive at the worst case \( ANB \) for the SHDA method, we assume that the shift takes place sometime between malformation \( z \) and malformation \( z + 1 \). In addition, let \( z' \) indicate the malformation that gives rise to the first flag after the shift occurs.

There are two conditions that will result in the worst possible \( ANB \) for the SHDA method. The first is similar to the Sets method: the sequence of births prior to the shift results in \( Z_0 \geq t_t \). This implies \( Z \geq t_t \).
which results in $G_{z+1} = 0$. The second condition is that the sequence of births prior to the shift results in $T_z \geq (b_t - n_t)$. Simply stated, the second condition means that no fewer than two flags must be raised in order for the chart to signal. As in Section 4.6, we define $Y_1$ as the number of births after the shift until (and including) the next malformation. Define $R_1$ as the number of births that occur after the first malformation following the shift until the next flag is raised, and $R_2$ as the number of births since the first flag was raised after the shift until a signal. The worst case $ANB$ for the SHDA method is then given by

$$W_t = E[Y_1] + E[R_1 \mid G_{z+1} = 0, T_{z+1} > (b_t - n_t)] + E[R_2 \mid G_{z'} = 0, T_{z'} = 0]$$

$$= \frac{1}{p_{1t}} + \frac{1-\theta_{n_t}^{n_t}}{p_{1t} \theta_{t_t}^{n_t}(1-\theta_{t_t})} + \frac{1-\theta_{n_t}^{n_t}}{p_{1t} \theta_{t_t}^{n_t}(1-\theta_{t_t})}$$

$$\Psi$$

$$= \frac{1}{p_{1t}} \left[ 1 + \frac{1-\theta_{n_t}^{n_t}}{\theta_{t_t}^{n_t}(1-\theta_{t_t})} \left( 1 + \frac{1}{\Psi} \right) \right]$$

where $\Psi$ is given by equation (4.8) and $\theta_t$ is calculated under incidence rate $p_1$. Note that the second term in the summation of the first line of equation (5.3) is equal to initial state $ANB_s(p_1)$ for the Sets method with $n_s = n_t$ and $t_s = t_t$. The third term is the initial state $ANB_t(p_1)$ for the SHDA method.

To identify the optimal design parameter values for $n_t$, $t_t$, and $b_t$, we use a grid search similar to the one described above for the Sets method. The search proceeds according to the following algorithm:

1. Begin with $b_t = 0$.  

Figure 5.1: Finding optimal $(n_s, t_s)$ for the Sets method when $p_0 = 0.001$, $\gamma_1 = 2$, and $m_0 = 50$. The solid points in the plots indicate the optimal values of $n_s$, $t_s$, $ANM_s(p_1)$, and the value of $ANM_s(p_0)$ that was achieved for $n_s = 5$ and $t_s = 747$. 

![Diagram A](A)

![Diagram B](B)

Minimum $ANM_s(0.002)$ occurs when $n_s = 5$ and $t_s = 747$. 

$$W_t = E[Y_1] + E[R_1 \mid G_{z+1} = 0, T_{z+1} > (b_t - n_t)] + E[R_2 \mid G_{z'} = 0, T_{z'} = 0]$$

$$= \frac{1}{p_{1t}} + \frac{1-\theta_{n_t}^{n_t}}{p_{1t} \theta_{n_t}^{n_t}(1-\theta_{n_t})} + \frac{1-\theta_{n_t}^{n_t}}{p_{1t} \theta_{n_t}^{n_t}(1-\theta_{n_t})}$$

$$\Psi$$

$$= \frac{1}{p_{1t}} \left[ 1 + \frac{1-\theta_{n_t}^{n_t}}{\theta_{n_t}^{n_t}(1-\theta_{n_t})} \left( 1 + \frac{1}{\Psi} \right) \right]$$

where $\Psi$ is given by equation (4.8) and $\theta_t$ is calculated under incidence rate $p_1$. Note that the second term in the summation of the first line of equation (5.3) is equal to initial state $ANB_s(p_1)$ for the Sets method with $n_s = n_t$ and $t_s = t_t$. The third term is the initial state $ANB_t(p_1)$ for the SHDA method.

To identify the optimal design parameter values for $n_t$, $t_t$, and $b_t$, we use a grid search similar to the one described above for the Sets method. The search proceeds according to the following algorithm:

1. Begin with $b_t = 0$.  

![Diagram A](A)

![Diagram B](B)

Minimum $ANM_s(0.002)$ occurs when $n_s = 5$ and $t_s = 747$. 

$$W_t = E[Y_1] + E[R_1 \mid G_{z+1} = 0, T_{z+1} > (b_t - n_t)] + E[R_2 \mid G_{z'} = 0, T_{z'} = 0]$$

$$= \frac{1}{p_{1t}} + \frac{1-\theta_{n_t}^{n_t}}{p_{1t} \theta_{n_t}^{n_t}(1-\theta_{n_t})} + \frac{1-\theta_{n_t}^{n_t}}{p_{1t} \theta_{n_t}^{n_t}(1-\theta_{n_t})}$$

$$\Psi$$

$$= \frac{1}{p_{1t}} \left[ 1 + \frac{1-\theta_{n_t}^{n_t}}{\theta_{n_t}^{n_t}(1-\theta_{n_t})} \left( 1 + \frac{1}{\Psi} \right) \right]$$

where $\Psi$ is given by equation (4.8) and $\theta_t$ is calculated under incidence rate $p_1$. Note that the second term in the summation of the first line of equation (5.3) is equal to initial state $ANB_s(p_1)$ for the Sets method with $n_s = n_t$ and $t_s = t_t$. The third term is the initial state $ANB_t(p_1)$ for the SHDA method.

To identify the optimal design parameter values for $n_t$, $t_t$, and $b_t$, we use a grid search similar to the one described above for the Sets method. The search proceeds according to the following algorithm:

1. Begin with $b_t = 0$.
2. Let \( b_t = b_t + 1 \) and \( n_t = 0 \).

3. Let \( n_t = n_t + 1 \). Note that for fixed \( n_t \), \( b_t \) and \( p \), \( ANM_t(p) \) is a decreasing function of \( t_t \). Using equation (4.7), calculate \( ANM_t(p_0) \) for various integer values of \( t_t \) until a value of \( t_t \) is found such that \( ANM_t(p_0) \geq m_0 \) but \( t_t + 1 \) results in \( ANM_t(p_0) < m_0 \).

4. Using equation (5.3), calculate and record \( W_t \) using the values of \( n_t \) and \( t_t \) found in step (3).

5. Return to step (3) and repeat until \( n_t = b_t \) or until it is clear that incrementing the value of \( n_t \) only results in increasing values of \( W_t \).

6. Record the minimal value of \( W_t \) that resulted from the current value of \( b_t \). Return to step (2) and repeat until these minimal values of \( W_t \) show an increasing pattern or until a prespecified limit of \( b_t \) is reached. (In our study, our limit of \( b_t = 150 \) was more than ample).

7. Choose the \((n_t, b_t, t_t)\) combination that results in the smallest \( W_t \).

Figure 5.2: Surface of \( W_t \) for various combinations of \( n_t \) and \( b_t \) for \( p_0 = 0.001 \) and \( m_0 = 200 \) \((b_0 = 200000)\) to detect a shift of size \( \gamma_1 = 3 \). In this case the minimum \( W_t \) occurs when \( n_t = 3 \) and \( b_t = 3 \).
Figure 5.2 gives a visual representation of the surface of \( W_t \) as a function of \( n_t \) and \( b_t \), subject to \( ANM_t(p_0) \geq 200 \) for detecting a shift of size \( \gamma_1 = 3 \) when \( p_0 = 0.001 \). The contours only occupy the lower right portion of the plot because \( n_t \) can be no larger than \( b_t \). The contours in Figure 5.2 show that for a fixed value of \( b_t \), \( W_t \) appears to be a convex function of \( n_t \). This contour pattern was typical among the combinations of \( p_0 \), \( m_0 \), and \( \gamma_1 \) that were used to compare the methods in Chapter 6.

5.4 CUSCORE method

Implementation of the CUSCORE method requires the specification of \( n_c \) and \( t_c \). Just as we did for the Sets and SHDA methods, to determine the worst case \( ANB \) for the CUSCORE method, we assume that the shift takes place sometime between the malformations \( z \) and \( z + 1 \). The worst possible \( ANB \) for the CUSCORE method occurs when the sequence of births prior to the shift results in \( C_z = 0 \) or \( C_z = 1 \) and \( Z_0 \geq t_c \). This forces \( C_{z+1} \) to reset to 0 and thus the first malformation observed after the shift does not contribute to the evidence that the shift has taken place. Once again, we define \( Y_1 \) as the number of births after the shift until (and including) the next malformation. Let \( R \) be the number of births that occur after the first malformation following the shift until the chart signals. Then the worst case \( ANB \) for the CUSCORE method is given by

\[
W_c = E(Y_1) + E(R \mid C_{z+1} = 0) = \frac{1}{p_1} [1 + ANM_c(p_1)]
\]

where \( ANM_c(p_1) \) is given by equation (4.12). Since the only portion of \( W_c \) that depends on the design parameters \( n_c \) and \( t_c \) is \( ANM_c(p_1) \), the values of \( n_c \) and \( t_c \) that minimize \( W_c \) are equivalent to those that minimize \( ANM_c(p_1) \).

Radaelli [19] gave an algorithm for choosing the values of \( n_c \) and \( t_c \) that minimize \( ANM_c(p_1) \) subject to \( ANM_c(p_0) \geq m_0 \). However, for consistency with respect to the other methods we employed a grid search to identify the optimal values for \( n_c \) and \( t_c \) given the values of \( p_0 \), \( \gamma_1 \), and \( m_0 \). The algorithm for the grid search is identical to the one shown for the Sets method in Section 5.2, except replace all \( s \) subscripts with \( c \) subscripts, and use equation (4.12) for the calculation of \( ANM_c \). As with the Sets method, we found that \( ANM_c(p_1) \), subject to \( ANM_c(p_0) \geq m_0 \), while not smooth, is a convex shaped function of \( n_c \). This concurs with results from Radaelli [19] who found evidence “that the absolute minimum of \( ANM_c(p_1) \) with respect to \( [n_c] \) is the first minimum on \( [n_c] \)”.

46
5.5 Bernoulli CUSUM

The design parameters of the Bernoulli CUSUM are $\delta$ and $h$. Moustakides’ result states that if $\delta$ is given by the critical inequality of the sequential probability ratio test, the CUSUM will be optimal. This value of $\delta$ is given in equation (3.5). If we permit a little flexibility in our choice of $\gamma_1$, we can adjust $\gamma_1$ slightly to $\gamma_1a$ and arrive at a value for $\delta$ that is the reciprocal of a positive integer, $m$. The result is that the CUSUM statistic $B_j$ only takes on a finite number of values which permits the exact calculation of the $ANB$ using Markov chains (see Section 3.4). Reynolds and Stoumbos [11] gave an algorithm for finding a suitable value of $\gamma_1a$ that is closest to the original choice of $\gamma_1$. Once $\delta = 1/m$ is identified, the value of $h$ is chosen so that $ANB_b(p_0) \geq b_0$. This can be accomplished by using the explicit solutions for $ANB_b$ described in Section 4.8, coupled with the fact that $ANB_b(p_0)$ is a non-decreasing function of $h$. 
Chapter 6

Comparison of methods

6.1 Design of the study

Our objective is to compare the Sets, SHDA, CUSCORE, and Bernoulli CUSUM methods over a reasonably broad range of combinations of the baseline incidence rate, \( p_0 \), the step-shift for which the charts are optimized, \( \gamma_1 \), and the target in-control average number of malformations until signal, \( m_0 \). Because the Sets, SHDA, and CUSCORE methods are designed to monitor geometric random variables, it was more convenient to pick target values of the in-control average run length in the scale of the average number of malformations until signal, as opposed to the average number of births until signal. For each of the methods, \( ANB(p_0) = ANM(p_0)/p_0 \) which implies that \( b_0 \), the target in-control average number of births until signal, must be at least as large as \( 1/p_0 \), or equivalently, \( m_0 \) must be greater than or equal to unity. For the sets based methods, \( m_0 \) needs to be appreciably larger than one in order for the optimal parameter choices of \( n_s, n_c, \) and \( n_t \) to be larger than one. Otherwise, the Sets and CUSCORE methods reduce to Shewhart charts—and the SHDA method with \( n_t = 1 \) is also similar to a Shewhart chart. As an example, suppose \( p_0 = 0.001 \) and we want to quickly detect a doubling of the incidence rate (\( \gamma_1 = 2 \)). Then for the Sets and CUSCORE methods, \( m_0 \) must be at least 6 in order for the optimal choices of \( n_s \) and \( n_c \) to be greater than one. For the SHDA method, \( m_0 \) must be at least 13 in order for the optimal \( n_t \) to be larger than one. This issue, coupled with the fact that we want the false alarm rate to be reasonably low, led us to consider values of \( m_0 \) ranging from 50 to 1000.

To make the comparisons between the methods as fair as possible, we searched for combinations of \( p_0, \gamma_1, \) and \( m_0 \) that resulted in the \( ANB(p_0) \) of the four methods being as close as possible. We then examined
the steady-state average number of births from shift until signal (\(SSANB\)) over a range of hypothetical shifts. Of course, if a shift in the incidence rate does occur, it is most likely to be something other than the shift for which the chart was designed, which is why we evaluated chart performance at shifts over a range of possible shifts.

Since the four methods we consider are based on discrete random variables, it is impossible to match the \(ANB(p_0)\) of all four methods exactly. In order to identify the combinations of \(p_0, \gamma_1,\) and \(m_0\) that gave the best matches of the \(ANB(p_0)\), we considered 7,600 combinations of \(p_0, \gamma_1,\) and \(m_0\), and, for each combination, we used the methodology discussed in Chapter 5 to identify the optimal choices of parameter values for each of the four methods. For each combination, the CUSUM parameters were chosen first, since that required a small adjustment of \(\gamma_1\) to \(\gamma_1a\). The optimal parameters for the other three methods were then chosen to detect a shift of size \(\gamma_1a\).

The 7,600 combinations of \(p_0, \gamma_1,\) and \(m_0\) were created as follows. Nineteen values of \(p_0\) were used, ranging from 0.01 to 0.0001. These were 0.01, 0.009, 0.008, 0.007, 0.006, 0.005, 0.004, 0.003, 0.002, 0.001, 0.0009, 0.0008, 0.0007, 0.0006, 0.0005, 0.0004, 0.0003, 0.0002, and 0.0001. The values of \(p_0\) are more concentrated near 0.0001 to focus on rare events, typical of the incidence rates of congenital malformations [24]. Twenty values of \(\gamma_1\) were used, ranging from 1.25 to 6 in increments of 0.25. Twenty values of \(m_0\) were used, ranging from 50 to 1000 in increments of 50. After the optimal parameter choices were determined for the four methods for each of the 7,600 combinations, 32 combinations were selected for comparison using the \(SSANB\) to measure the performance across a range of hypothetical shifts. These 32 combinations were selected according to the following criteria:

1. Only combinations where the CUSUM had the highest \(ANB(p_0)\) were considered. That is, where \(ANB_b(p_0) \geq \max[ANB_s(p_0), ANB_t(p_0), ANB_c(p_0)]\). Since we anticipate the steady-state performance of the CUSUM to be better than the other methods, we wished to ensure that it did not have an advantage over the other three methods by having a smaller \(ANB(p_0)\).

2. The standard deviation of \(ANB_s(p_0), ANB_t(p_0), ANB_c(p_0),\) and \(ANB_b(p_0)\) was used as a simple metric to ascertain how close together the \(ANB(p_0)\) values were for the four methods. Only the 3,218 (42%) combinations for which the four \(ANB(p_0)\) values had standard deviations less than 1,000 were considered.

3. Seventy-one combinations met the two criterion stated above. From this group of 71, 32 combinations were selected to give as much coverage as possible over the three-dimensional space generated by the
selected combinations of \( p_0, \gamma_1, \) and \( m_0 \). Combinations that had four \( ANB(p_0) \) values with smaller standard deviations were given preference when possible.

We noticed that the selection of the 32 combinations above did not result in any combination with a value of \( \gamma_1 \) less than 2.5. In order to assess the impact of designing the chart for a small shift, we chose to relax the first criterion stated above and also consider combinations of \( p_0, \gamma_1, \) and \( m_0 \) that satisfied the following criteria: 1) the CUSUM had the second largest \( ANB(p_0) \) among the four methods, 2) the standard deviation of the \( ANB(p_0) \) for the four methods was less than 100, and 3) \( \gamma_1 < 2.5 \). There were 13 such combinations that satisfied these criteria, 8 of which were selected to be included with the 32 original combinations. Thus, the \( SSANB \) performance was calculated for a total of 40 combinations, or “cases.” Figure 6.1 shows the coverage of \( p_0, \gamma_1, \) and \( m_0 \) that was achieved by the 40 cases. Each point in the figure represents a case for which the \( SSANB \) of the four methods were compared over a range of hypothetical shifts.

Table 6.1 gives the optimal parameter values for the four methods for each of the 40 cases of \( p_0, \gamma_1, \) and \( m_0 \). Of particular interest is that for each of the 40 cases, \( n_t = b_t \) for the SHDA method. In fact, only three of the 7,600 combinations that were originally considered did not result in \( n_t = b_t \). Recall that if \( n_t = b_t \) and if the first \( n_t \) sets do not fall below the threshold \( t_t \), the SHDA method is identical to the Sets method with \( n_s = 2n_t \). Also note in Table 6.1 that \( n_t \) is roughly half of \( n_s \). Hence, when the Sets and SHDA methods are designed to be optimal in the sense set forth by Lorden [37] and Moustakides [38], the two methods bear a striking resemblance. It is important to note why the choices for \( n_t \) and \( b_t \) shown in Table 6.1 differ considerably from those given in Table 1' and Table 1'' of Sitter et al. [23]. The optimality criteria used by Sitter et al. was to chose the parameters that minimized \( ANM_t(p_1) \) subject to \( ANM_t(p_0) \geq m_0 \), whereas in this work, the parameters were chosen to minimize \( W_t \) (see equation (5.3)) subject to \( ANM_t(p_0) \geq m_0 \). The implications of this are discussed in Section 7.3.1.

Table 6.2 shows the values of \( ANB(p_0) \) that were achieved for the cases considered. Note that for each case, \( ANB_6(p_0) \) (the average in-control number of births until signal for the CUSUM) is the highest among the four methods except cases 1, 5, 8, 9, 12, 15, 22, and 36 where \( ANB_6(p_0) \) is second highest. Thus, in most of the cases, the CUSUM chart had a slight disadvantage compared to the sets based methods because it had the largest in-control average run length, or, in other words, the lowest false alarm rate.
Figure 6.1: Forty cases of $p_0$, $\gamma_1$, and $m_0$ that were used to compare the four methods with respect to SSANB performance.

6.2 Comparison of methods using SSANB

To evaluate the steady-state average run length performance across a range of possible shifts, let $\gamma$ denote a hypothetical shift for which SSANB will be calculated. For each of the 40 combinations, we calculated SSANB($\gamma$) for 29 values of $\gamma = \{\gamma_{1a}, 1.25, 1.5, \ldots, 7.75, 8\}$. Note that $\gamma_{1a}$ is the precise shift that the charts were optimally designed to detect. For the Sets, SHDA, and CUSCORE methods, the SSANB calculations were made exactly using equations (4.6), (4.11), and (4.14). For the CUSUM, the SSANB was calculated exactly using equation (4.23), except for the six cases (34, 36-40) where the number of states in the Markov chain exceeded 5,000. In these cases, the algorithm described in Section 4.9 was used to generate one million Monte Carlo replications to estimate the SSANB. In the process of comparing the four methods, these $6 \times 29 = 174$ simulated estimates of $SSANB_6(\gamma)$ were compared to the corresponding minimum value of $SSANB_a(\gamma)$, $SSANB_b(\gamma)$, and $SSANB_c(\gamma)$. To control the overall type I error of these 174 comparisons,
<table>
<thead>
<tr>
<th>Case</th>
<th>$p_0$</th>
<th>$\gamma_1$</th>
<th>$m_0$</th>
<th>$\gamma_1^c$</th>
<th>$n_x$</th>
<th>$t_x$</th>
<th>$n_t$</th>
<th>$b_t$</th>
<th>$t_t$</th>
<th>$n_c$</th>
<th>$t_c$</th>
<th>$m$</th>
<th>$h$</th>
<th>$w$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0100</td>
<td>1.50</td>
<td>150</td>
<td>1.50274</td>
<td>14</td>
<td>149</td>
<td>6</td>
<td>6</td>
<td>128</td>
<td>9</td>
<td>62</td>
<td>81</td>
<td>6.74074</td>
<td>546</td>
</tr>
<tr>
<td>2</td>
<td>0.0100</td>
<td>4.00</td>
<td>900</td>
<td>4.00717</td>
<td>7</td>
<td>52</td>
<td>4</td>
<td>4</td>
<td>61</td>
<td>7</td>
<td>39</td>
<td>46</td>
<td>4.84783</td>
<td>223</td>
</tr>
<tr>
<td>3</td>
<td>0.0090</td>
<td>3.75</td>
<td>100</td>
<td>3.78005</td>
<td>5</td>
<td>65</td>
<td>2</td>
<td>2</td>
<td>46</td>
<td>4</td>
<td>41</td>
<td>53</td>
<td>3.41509</td>
<td>181</td>
</tr>
<tr>
<td>4</td>
<td>0.0090</td>
<td>4.75</td>
<td>200</td>
<td>4.74481</td>
<td>5</td>
<td>53</td>
<td>3</td>
<td>3</td>
<td>66</td>
<td>4</td>
<td>33</td>
<td>46</td>
<td>3.54348</td>
<td>163</td>
</tr>
<tr>
<td>5</td>
<td>0.0080</td>
<td>2.00</td>
<td>50</td>
<td>1.98370</td>
<td>5</td>
<td>93</td>
<td>3</td>
<td>3</td>
<td>108</td>
<td>4</td>
<td>59</td>
<td>87</td>
<td>3.93103</td>
<td>342</td>
</tr>
<tr>
<td>6</td>
<td>0.0080</td>
<td>2.25</td>
<td>550</td>
<td>2.25186</td>
<td>13</td>
<td>138</td>
<td>6</td>
<td>6</td>
<td>127</td>
<td>9</td>
<td>63</td>
<td>81</td>
<td>6.40741</td>
<td>519</td>
</tr>
<tr>
<td>7</td>
<td>0.0080</td>
<td>4.75</td>
<td>900</td>
<td>4.71240</td>
<td>7</td>
<td>65</td>
<td>3</td>
<td>3</td>
<td>52</td>
<td>6</td>
<td>42</td>
<td>52</td>
<td>4.50000</td>
<td>234</td>
</tr>
<tr>
<td>8</td>
<td>0.0070</td>
<td>1.25</td>
<td>100</td>
<td>1.24072</td>
<td>15</td>
<td>246</td>
<td>7</td>
<td>7</td>
<td>224</td>
<td>8</td>
<td>91</td>
<td>128</td>
<td>7.22656</td>
<td>925</td>
</tr>
<tr>
<td>9</td>
<td>0.0070</td>
<td>1.25</td>
<td>700</td>
<td>1.24072</td>
<td>35</td>
<td>303</td>
<td>18</td>
<td>18</td>
<td>304</td>
<td>20</td>
<td>94</td>
<td>128</td>
<td>13.83594</td>
<td>1771</td>
</tr>
<tr>
<td>10</td>
<td>0.0070</td>
<td>2.75</td>
<td>350</td>
<td>2.77838</td>
<td>9</td>
<td>121</td>
<td>4</td>
<td>4</td>
<td>105</td>
<td>6</td>
<td>58</td>
<td>82</td>
<td>5.10976</td>
<td>419</td>
</tr>
<tr>
<td>11</td>
<td>0.0070</td>
<td>3.75</td>
<td>100</td>
<td>3.70731</td>
<td>5</td>
<td>84</td>
<td>2</td>
<td>2</td>
<td>59</td>
<td>4</td>
<td>53</td>
<td>69</td>
<td>3.44928</td>
<td>238</td>
</tr>
<tr>
<td>12</td>
<td>0.0060</td>
<td>1.25</td>
<td>400</td>
<td>1.24609</td>
<td>27</td>
<td>328</td>
<td>12</td>
<td>12</td>
<td>300</td>
<td>15</td>
<td>108</td>
<td>149</td>
<td>11.63758</td>
<td>1734</td>
</tr>
<tr>
<td>13</td>
<td>0.0060</td>
<td>2.25</td>
<td>350</td>
<td>2.25254</td>
<td>10</td>
<td>157</td>
<td>6</td>
<td>6</td>
<td>183</td>
<td>8</td>
<td>84</td>
<td>108</td>
<td>5.87963</td>
<td>635</td>
</tr>
<tr>
<td>14</td>
<td>0.0060</td>
<td>3.75</td>
<td>100</td>
<td>3.74716</td>
<td>5</td>
<td>98</td>
<td>2</td>
<td>2</td>
<td>69</td>
<td>4</td>
<td>62</td>
<td>80</td>
<td>3.43750</td>
<td>275</td>
</tr>
<tr>
<td>15</td>
<td>0.0050</td>
<td>2.00</td>
<td>50</td>
<td>1.98943</td>
<td>5</td>
<td>149</td>
<td>3</td>
<td>3</td>
<td>173</td>
<td>5</td>
<td>115</td>
<td>139</td>
<td>3.93525</td>
<td>547</td>
</tr>
<tr>
<td>16</td>
<td>0.0050</td>
<td>2.50</td>
<td>400</td>
<td>2.50364</td>
<td>9</td>
<td>165</td>
<td>5</td>
<td>5</td>
<td>181</td>
<td>7</td>
<td>90</td>
<td>122</td>
<td>5.61475</td>
<td>685</td>
</tr>
<tr>
<td>17</td>
<td>0.0050</td>
<td>4.50</td>
<td>50</td>
<td>4.48019</td>
<td>3</td>
<td>72</td>
<td>2</td>
<td>2</td>
<td>105</td>
<td>3</td>
<td>67</td>
<td>86</td>
<td>2.75581</td>
<td>237</td>
</tr>
<tr>
<td>18</td>
<td>0.0050</td>
<td>5.25</td>
<td>50</td>
<td>5.23016</td>
<td>3</td>
<td>72</td>
<td>2</td>
<td>2</td>
<td>105</td>
<td>3</td>
<td>67</td>
<td>78</td>
<td>2.62821</td>
<td>205</td>
</tr>
<tr>
<td>19</td>
<td>0.0040</td>
<td>5.00</td>
<td>750</td>
<td>5.03011</td>
<td>6</td>
<td>110</td>
<td>3</td>
<td>3</td>
<td>109</td>
<td>5</td>
<td>71</td>
<td>100</td>
<td>4.29000</td>
<td>429</td>
</tr>
<tr>
<td>20</td>
<td>0.0040</td>
<td>5.50</td>
<td>850</td>
<td>5.54341</td>
<td>6</td>
<td>107</td>
<td>3</td>
<td>3</td>
<td>106</td>
<td>5</td>
<td>69</td>
<td>94</td>
<td>4.19149</td>
<td>394</td>
</tr>
<tr>
<td>21</td>
<td>0.0040</td>
<td>5.75</td>
<td>750</td>
<td>5.73228</td>
<td>6</td>
<td>110</td>
<td>3</td>
<td>3</td>
<td>109</td>
<td>5</td>
<td>71</td>
<td>92</td>
<td>4.06522</td>
<td>374</td>
</tr>
<tr>
<td>22</td>
<td>0.0030</td>
<td>1.75</td>
<td>50</td>
<td>1.74604</td>
<td>6</td>
<td>307</td>
<td>3</td>
<td>3</td>
<td>289</td>
<td>5</td>
<td>192</td>
<td>249</td>
<td>4.26104</td>
<td>1061</td>
</tr>
<tr>
<td>23</td>
<td>0.0030</td>
<td>5.50</td>
<td>150</td>
<td>5.50304</td>
<td>4</td>
<td>125</td>
<td>2</td>
<td>2</td>
<td>121</td>
<td>4</td>
<td>109</td>
<td>126</td>
<td>3.21429</td>
<td>405</td>
</tr>
<tr>
<td>24</td>
<td>0.0020</td>
<td>5.75</td>
<td>200</td>
<td>5.74425</td>
<td>4</td>
<td>170</td>
<td>2</td>
<td>2</td>
<td>166</td>
<td>4</td>
<td>150</td>
<td>184</td>
<td>3.33696</td>
<td>614</td>
</tr>
<tr>
<td>25</td>
<td>0.0020</td>
<td>6.00</td>
<td>800</td>
<td>5.99585</td>
<td>6</td>
<td>217</td>
<td>3</td>
<td>3</td>
<td>215</td>
<td>5</td>
<td>140</td>
<td>179</td>
<td>4.04469</td>
<td>724</td>
</tr>
<tr>
<td>26</td>
<td>0.0010</td>
<td>3.25</td>
<td>100</td>
<td>3.24723</td>
<td>5</td>
<td>591</td>
<td>3</td>
<td>3</td>
<td>714</td>
<td>4</td>
<td>373</td>
<td>524</td>
<td>3.67557</td>
<td>1926</td>
</tr>
<tr>
<td>27</td>
<td>0.0010</td>
<td>6.00</td>
<td>150</td>
<td>6.00253</td>
<td>4</td>
<td>375</td>
<td>2</td>
<td>2</td>
<td>364</td>
<td>4</td>
<td>328</td>
<td>358</td>
<td>3.12291</td>
<td>1118</td>
</tr>
<tr>
<td>28</td>
<td>0.0009</td>
<td>3.50</td>
<td>850</td>
<td>3.49641</td>
<td>9</td>
<td>796</td>
<td>4</td>
<td>4</td>
<td>689</td>
<td>7</td>
<td>440</td>
<td>557</td>
<td>5.23399</td>
<td>2915</td>
</tr>
<tr>
<td>29</td>
<td>0.0009</td>
<td>4.50</td>
<td>50</td>
<td>4.50460</td>
<td>3</td>
<td>401</td>
<td>2</td>
<td>2</td>
<td>585</td>
<td>3</td>
<td>373</td>
<td>477</td>
<td>2.76520</td>
<td>1319</td>
</tr>
<tr>
<td>30</td>
<td>0.0007</td>
<td>4.50</td>
<td>350</td>
<td>4.49678</td>
<td>6</td>
<td>756</td>
<td>3</td>
<td>3</td>
<td>744</td>
<td>5</td>
<td>485</td>
<td>614</td>
<td>4.03094</td>
<td>2475</td>
</tr>
</tbody>
</table>

Table 6.1: Cases of $p_0$, $\gamma_1$, and $m_0$ with optimal design parameters ($w$ is the number of transient states in the Markov chain for the CUSUM).
<table>
<thead>
<tr>
<th>Case</th>
<th>$p_0$</th>
<th>$\gamma_1$</th>
<th>$m_0$</th>
<th>$b_0$</th>
<th>Sets</th>
<th>SHDA</th>
<th>CUSCORE</th>
<th>CUSUM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ANB($p_0$)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.0100</td>
<td>1.50</td>
<td>150</td>
<td>15000</td>
<td>15036</td>
<td>15006</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.0100</td>
<td>4.00</td>
<td>900</td>
<td>90000</td>
<td>90940</td>
<td>90645</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.0090</td>
<td>3.75</td>
<td>100</td>
<td>11111</td>
<td>11341</td>
<td>11113</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.0090</td>
<td>4.75</td>
<td>700</td>
<td>11368</td>
<td>11390</td>
<td>90728</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.0080</td>
<td>2.00</td>
<td>6250</td>
<td>6288</td>
<td>6288</td>
<td>6288</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.0080</td>
<td>2.25</td>
<td>550</td>
<td>68750</td>
<td>68796</td>
<td>69006</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0.0080</td>
<td>4.75</td>
<td>900</td>
<td>112500</td>
<td>114228</td>
<td>115057</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0.0070</td>
<td>1.25</td>
<td>14286</td>
<td>14310</td>
<td>14314</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0.0070</td>
<td>1.25</td>
<td>100000</td>
<td>100069</td>
<td>100066</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.0070</td>
<td>1.25</td>
<td>700</td>
<td>11341</td>
<td>11390</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>0.0070</td>
<td>3.75</td>
<td>100</td>
<td>14394</td>
<td>14436</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>0.0060</td>
<td>1.25</td>
<td>66667</td>
<td>66847</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>0.0060</td>
<td>2.75</td>
<td>58333</td>
<td>58447</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>0.0060</td>
<td>3.75</td>
<td>100000</td>
<td>100069</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>0.0050</td>
<td>2.00</td>
<td>100000</td>
<td>100069</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>0.0050</td>
<td>2.50</td>
<td>80000</td>
<td>80000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>0.0050</td>
<td>4.50</td>
<td>100000</td>
<td>100069</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>0.0050</td>
<td>5.25</td>
<td>100000</td>
<td>100069</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>0.0040</td>
<td>5.00</td>
<td>187500</td>
<td>188769</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>0.0040</td>
<td>5.50</td>
<td>212500</td>
<td>212981</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>0.0040</td>
<td>5.75</td>
<td>187500</td>
<td>188769</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>0.0030</td>
<td>1.75</td>
<td>16667</td>
<td>16701</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>0.0030</td>
<td>5.50</td>
<td>50000</td>
<td>50011</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>0.0020</td>
<td>5.75</td>
<td>200000</td>
<td>200069</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>0.0020</td>
<td>6.00</td>
<td>400000</td>
<td>400049</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>0.0010</td>
<td>3.25</td>
<td>100000</td>
<td>100000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>0.0010</td>
<td>6.00</td>
<td>150000</td>
<td>150000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>0.0009</td>
<td>3.50</td>
<td>850</td>
<td>850</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>0.0009</td>
<td>4.50</td>
<td>50</td>
<td>50</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>0.0007</td>
<td>4.50</td>
<td>350</td>
<td>350</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>0.0006</td>
<td>3.00</td>
<td>250</td>
<td>250</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>0.0005</td>
<td>3.50</td>
<td>100000</td>
<td>100000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>0.0005</td>
<td>5.75</td>
<td>50</td>
<td>50</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>0.0004</td>
<td>3.75</td>
<td>600</td>
<td>600</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>0.0004</td>
<td>4.75</td>
<td>250</td>
<td>250</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>0.0004</td>
<td>1.75</td>
<td>100</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>0.0003</td>
<td>2.25</td>
<td>600</td>
<td>600</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>0.0002</td>
<td>4.00</td>
<td>750</td>
<td>750</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>0.0002</td>
<td>5.25</td>
<td>100</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>0.0002</td>
<td>5.75</td>
<td>350</td>
<td>350</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 6.2: Values of $ANB(p_0)$ that were achieved for the 40 cases of $p_0$, $\gamma_1$, and $m_0$.

95% simultaneous Bonferroni confidence intervals were calculated for the $SSANB_b$ using the standard normal quantile $Z_{1-\alpha/2}\cdot \sqrt{\frac{b}{2}} = 3.63$ as the critical value.
Table 6.3 shows the steady-state average number of births from shift until signal when the shift was equal to the shift for which the chart was designed ($SSANB(\gamma_{1a})$). In all forty cases, the CUSUM has the lowest $SSANB(\gamma_{1a})$. For the six simulated cases, note that the upper limit of the 95% simultaneous confidence interval for $SSANB_0(\gamma_{1a})$ is still lower than the $SSANB$ of the other three methods. Another important feature in Table 6.3 is that with the exception of cases 2, 32, and 39, the following inequality holds:

\[ SSANB_b(\gamma_{1a}) < SSANB_c(\gamma_{1a}) < SSANB_s(\gamma_{1a}) < SSANB_t(\gamma_{1a}) \]  

6.1

Examination of case 2 in Table 6.3 reveals that the SHDA method outperforms the Sets method. In cases 32 and 39, the Sets method has better $SSANB$ performance than the CUSCORE method. These three deviations from the inequality in (6.1) may be explained in part by the fact that $ANB_t(p_0) < ANB_s(p_0)$ for case 2 and $ANB_s(p_0) < ANB_c(p_0)$ in case 32. (refer to Table 6.2). We reemphasize that the $SSANB$ performance of the CUSUM was best at the shift for which the chart was designed in all forty cases.

The inequality in equation (6.1) demonstrates the efficiency with which the four methods accumulate information and react to a shift in the incidence rate. Namely, the Bernoulli CUSUM is best, followed by the CUSCORE, then the Sets method, and last of all, the SHDA method. The SHDA method is worst because of the possibility that at least two flags must be raised in order to signal—thereby incurring extra delay in detecting the shift. Although the delay in the Sets method is not as pronounced, requiring that $n_s$ sets in a row be less than the threshold forces the counter $S_t$ to reset whenever a set exceeds the threshold $t_s$, thereby discarding all previous information accumulated in the chart. Among the sets based methods, the CUSCORE is the most efficient because a set that exceeds the threshold does not necessarily reset the chart, but rather moves it down one step, allowing the chart to accumulate information over a longer period of time. But ultimately, the dichotomization of sets into those that exceed the threshold and those that do not reduces the amount of information contained in the chart, which is why the CUSUM is the most efficient of the four methods we considered.

The comparison of the methods across a range of possible shifts is best done graphically. Plots of the $SSANB$ profile (like those shown in Figures 6.2 and 6.3) were created for each of the 40 cases. (The complete set of plots is included in Appendix A). One of the most notable features of these plots was that for every case, the $SSANB$ performance of the CUSUM is almost always better than all the other methods across the entire range of hypothetical shifts that were considered. For each case and each method, $SSANB$ was
<table>
<thead>
<tr>
<th>Combinations</th>
<th>(SSANB(\gamma_{1a}))</th>
<th>Sets</th>
<th>SHDA</th>
<th>CUSCORE</th>
<th>CUSUM</th>
<th>95% Simul. C.I.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Case</strong></td>
<td><strong>(p_0)</strong></td>
<td><strong>(\gamma_1)</strong></td>
<td><strong>(m_0)</strong></td>
<td><strong>(\gamma_{1a})</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.0100</td>
<td>1.50</td>
<td>150</td>
<td>1.50274</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.0100</td>
<td>4.00</td>
<td>900</td>
<td>4.00717</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.0090</td>
<td>3.75</td>
<td>100</td>
<td>3.78005</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.0090</td>
<td>4.75</td>
<td>200</td>
<td>4.74481</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.0080</td>
<td>2.00</td>
<td>50</td>
<td>1.98370</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.0080</td>
<td>2.25</td>
<td>550</td>
<td>2.25186</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0.0070</td>
<td>1.25</td>
<td>100</td>
<td>1.24072</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0.0070</td>
<td>1.25</td>
<td>700</td>
<td>1.24072</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0.0070</td>
<td>2.75</td>
<td>350</td>
<td>2.77834</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.0070</td>
<td>3.75</td>
<td>100</td>
<td>3.70731</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>0.0060</td>
<td>1.25</td>
<td>400</td>
<td>1.24609</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>0.0060</td>
<td>4.00</td>
<td>900</td>
<td>4.00717</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>0.0050</td>
<td>2.00</td>
<td>50</td>
<td>1.98943</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>0.0050</td>
<td>2.50</td>
<td>50</td>
<td>2.50364</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>0.0050</td>
<td>5.25</td>
<td>50</td>
<td>5.23016</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>0.0050</td>
<td>4.50</td>
<td>50</td>
<td>4.48019</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>0.0050</td>
<td>3.75</td>
<td>100</td>
<td>3.74716</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>0.0040</td>
<td>1.25</td>
<td>400</td>
<td>1.24609</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>0.0040</td>
<td>5.50</td>
<td>150</td>
<td>5.50304</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>0.0040</td>
<td>5.75</td>
<td>750</td>
<td>5.73228</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>0.0040</td>
<td>7.50</td>
<td>50</td>
<td>7.53228</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>0.0030</td>
<td>1.75</td>
<td>50</td>
<td>1.74604</td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>0.0030</td>
<td>5.50</td>
<td>150</td>
<td>5.50304</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>0.0020</td>
<td>5.75</td>
<td>200</td>
<td>5.74425</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>0.0020</td>
<td>6.00</td>
<td>800</td>
<td>5.99585</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>0.0010</td>
<td>3.25</td>
<td>100</td>
<td>3.24723</td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>0.0010</td>
<td>6.00</td>
<td>150</td>
<td>6.00253</td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>0.0009</td>
<td>3.50</td>
<td>850</td>
<td>3.49641</td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>0.0009</td>
<td>4.50</td>
<td>50</td>
<td>4.50460</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>0.0007</td>
<td>4.50</td>
<td>350</td>
<td>4.49678</td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>0.0006</td>
<td>3.00</td>
<td>250</td>
<td>3.00229</td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>0.0005</td>
<td>3.50</td>
<td>50</td>
<td>3.50048</td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>0.0005</td>
<td>5.75</td>
<td>450</td>
<td>5.75320</td>
<td></td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>0.0004</td>
<td>3.75</td>
<td>600</td>
<td>3.75234</td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>0.0004</td>
<td>4.75</td>
<td>250</td>
<td>4.74684</td>
<td></td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>0.0004</td>
<td>1.75</td>
<td>100</td>
<td>1.75062</td>
<td></td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>0.0003</td>
<td>2.25</td>
<td>600</td>
<td>2.25078</td>
<td></td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>0.0002</td>
<td>4.00</td>
<td>750</td>
<td>4.00095</td>
<td></td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>0.0002</td>
<td>5.25</td>
<td>100</td>
<td>5.24846</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>0.0002</td>
<td>5.75</td>
<td>350</td>
<td>5.75010</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 6.3: Values of \(SSANB(\gamma_{1a})\) for the 40 cases of \(p_0\), \(\gamma_1\), and \(m_0\). The 95% simultaneous confidence interval is \(SSANB_b(\gamma_{1a}) \pm 3.63 \times SE\) for the six cases when the CUSUM was simulated.
calculated for 29 values of \( \gamma = \{\gamma_1, 1.25, 1.5, \ldots, 7.75, 8\} \). Thus, among the \( 40 \times 29 = 1160 \) comparisons of the four methods, the \( SSANB \) for the CUSUM (or the upper limit of the simultaneous confidence interval, if \( SSANB(\gamma) \) was simulated) was less than all three of the corresponding \( SSANB \) values of the sets based methods, except for 13 instances where the CUSCORE method had the best \( SSANB \) performance. These 13 instances are are shown in Table 6.4. Note that the relative improvement of the CUSCORE over the CUSUM is very small, and that this situation occurs when the chart is designed to detect a relatively large shift (\( \gamma_1 \geq 4 \)) and a small shift (\( \gamma \leq 1.75 \)) actually occurs.

![Figure 6.2: Comparing \( SSANB \) performance across a range of hypothetical shifts for cases 12 and 13.](image)

In addition to demonstrating the almost uniform dominance of the CUSUM over the other methods, a study of the \( SSANB \) profile graphs for the 40 cases revealed some other interesting patterns. In particular, differences between the methods appear to be most affected by \( \gamma_1 \), the target shift for which the charts were designed. Case 12 in Figure 6.2 shows a pattern that is typical when \( \gamma_1 < 2 \). Note how the \( SSANB \) profiles of the Sets and SHDA methods diverge rapidly from the CUSCORE and CUSUM for \( \gamma > 3 \), where the CUSUM and CUSCORE perform considerably better if a large shift is observed. This differentiation between the Sets and SHDA methods and the CUSCORE and CUSUM is clearly visible in cases 1, 8, 9, 12, 22, 36, and 37, which were designed for shifts of size 1.5, 1.25, 1.25, 1.25, 1.75, 1.75, and 2.25, respectively. Case 13 in Figure 6.2 is representative of shift sizes in the neighborhood of 2. Note how the \( SSANB \) of the methods are clearly distinguished from one another, and that the inequality in (6.1) appears to be preserved across the values of \( \gamma \). Cases whose profiles have a similar pattern to that of case 13 are cases 5, 15, 16, 26,
Case 28: $p_0 = 0.0009$, $\gamma_1 = 3.5$, $m_0 = 850$

Case 40: $p_0 = 0.0002$, $\gamma_1 = 5.75$, $m_0 = 350$

Figure 6.3: Comparing $SSANB$ performance across a range of hypothetical shifts for cases 28 and 40. Since the $SSANB_b$ is simulated for case 40, the width of the solid line for the CUSUM is given by $SSANB_b(\gamma) \pm 3.63 \times SE$.

<table>
<thead>
<tr>
<th>Case</th>
<th>$p_0$</th>
<th>$\gamma_1$</th>
<th>$m_0$</th>
<th>$\gamma$</th>
<th>$SSANB_c(\gamma)$</th>
<th>$SSANB_b(\gamma)$</th>
<th>Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.0100</td>
<td>4.00</td>
<td>900</td>
<td>1.25</td>
<td>20162.78</td>
<td>21889.94</td>
<td>1.086</td>
</tr>
<tr>
<td>2</td>
<td>0.0100</td>
<td>4.00</td>
<td>900</td>
<td>1.50</td>
<td>6941.18</td>
<td>7383.20</td>
<td>1.064</td>
</tr>
<tr>
<td>2</td>
<td>0.0100</td>
<td>4.00</td>
<td>900</td>
<td>1.75</td>
<td>3233.10</td>
<td>3242.17</td>
<td>1.003</td>
</tr>
<tr>
<td>7</td>
<td>0.0080</td>
<td>4.75</td>
<td>900</td>
<td>1.25</td>
<td>29533.68</td>
<td>30184.24</td>
<td>1.022</td>
</tr>
<tr>
<td>23</td>
<td>0.0030</td>
<td>5.50</td>
<td>150</td>
<td>1.25</td>
<td>20193.29</td>
<td>20378.71</td>
<td>1.009</td>
</tr>
<tr>
<td>27</td>
<td>0.0010</td>
<td>6.00</td>
<td>150</td>
<td>1.25</td>
<td>60259.36</td>
<td>62162.78</td>
<td>1.032</td>
</tr>
<tr>
<td>27</td>
<td>0.0010</td>
<td>6.00</td>
<td>150</td>
<td>1.50</td>
<td>30176.81</td>
<td>31151.18</td>
<td>1.032</td>
</tr>
<tr>
<td>27</td>
<td>0.0010</td>
<td>6.00</td>
<td>150</td>
<td>1.75</td>
<td>17567.29</td>
<td>17891.25</td>
<td>1.018</td>
</tr>
<tr>
<td>33</td>
<td>0.0005</td>
<td>5.75</td>
<td>450</td>
<td>1.25</td>
<td>287379.16</td>
<td>299605.40</td>
<td>1.043</td>
</tr>
<tr>
<td>33</td>
<td>0.0005</td>
<td>5.75</td>
<td>450</td>
<td>1.50</td>
<td>121683.03</td>
<td>126609.69</td>
<td>1.040</td>
</tr>
<tr>
<td>33</td>
<td>0.0005</td>
<td>5.75</td>
<td>450</td>
<td>1.75</td>
<td>62672.61</td>
<td>63717.80</td>
<td>1.017</td>
</tr>
<tr>
<td>35</td>
<td>0.0004</td>
<td>4.75</td>
<td>250</td>
<td>1.25</td>
<td>210640.14</td>
<td>217617.06</td>
<td>1.033</td>
</tr>
<tr>
<td>35</td>
<td>0.0004</td>
<td>4.75</td>
<td>250</td>
<td>1.50</td>
<td>94706.96</td>
<td>96745.21</td>
<td>1.022</td>
</tr>
</tbody>
</table>

Table 6.4: Thirteen instances when the CUSCORE had better $SSANB$ performance than the CUSUM. The Ratio is $SSANB_b(\gamma)/SSANB_c(\gamma)$.

and 31, which have corresponding values of $\gamma_1$ equal to 2, 2, 2.5, 3.25, and 3. Cases 28 and 40 in Figure 6.3 are typical examples of profiles from cases with relatively large shifts ($\gamma_1 \geq 3$). In these cases it is more difficult to distinguish between the Sets, SHDA, and CUSCORE methods, and the inequality in (6.1) is not preserved across the range of $\gamma$. Table 6.5 gives the actual $SSANB$ values for cases 12, 13, 28, and 40 for selected values of $\gamma$. If desired, the $SSANB$ can easily be converted to the steady-state average time from
shift until signal by dividing the $SSANB$ by the birth rate.

$$SSANB(\gamma)$$  | Sets | SHDA | CUSCORE | CUSUM | 95% Simul. C.I. \\
--- | --- | --- | --- | --- | --- \\
12 | 1.5 | 6200.15 | 6508.67 | 4916.07 | 3771.39 \\
12 | 2.0 | 2585.77 | 2544.21 | 2090.25 | 1713.17 \\
12 | 2.5 | 1689.56 | 1600.89 | 1274.46 | 1104.67 \\
12 | 3.0 | 1306.42 | 1213.15 | 907.14 | 814.98 \\
12 | 4.0 | 940.52 | 860.17 | 576.20 | 534.75 \\
12 | 6.0 | 620.30 | 564.83 | 342.24 | 317.10 \\
12 | 8.0 | 464.07 | 422.40 | 248.96 | 225.47 \\
13 | 1.5 | 6773.99 | 6688.15 | 5455.86 | 4604.99 \\
13 | 2.0 | 2232.63 | 2232.22 | 1910.45 | 1517.49 \\
13 | 2.5 | 1164.46 | 1202.19 | 1053.70 | 846.65 \\
13 | 3.0 | 764.78 | 814.11 | 706.75 | 581.11 \\
13 | 4.0 | 459.38 | 510.96 | 418.47 | 355.83 \\
13 | 6.0 | 272.75 | 314.52 | 232.21 | 200.34 \\
13 | 8.0 | 208.47 | 233.07 | 164.05 | 139.50 \\
28 | 1.5 | 90485.66 | 101734.43 | 74333.57 | 68679.24 \\
28 | 2.0 | 24689.58 | 27617.52 | 20161.09 | 16721.01 \\
28 | 2.5 | 11120.60 | 12109.52 | 9561.94 | 7646.74 \\
28 | 3.0 | 6554.58 | 6926.11 | 5874.29 | 4710.29 \\
28 | 4.0 | 3428.46 | 3442.88 | 3157.05 | 2607.00 \\
28 | 6.0 | 1808.52 | 1711.58 | 1601.57 | 1368.23 \\
28 | 8.0 | 1283.43 | 1184.82 | 1084.61 | 927.47 \\
40 | 1.5 | 297980.88 | 279738.05 | 310743.45 | 268576.41 (267608.6, 269544.2) \\
40 | 2.0 | 99484.16 | 90205.51 | 103943.78 | 82054.44 (81768.0, 82340.9) \\
40 | 2.5 | 47159.69 | 42856.26 | 48999.60 | 37322.63 (37198.4, 37446.8) \\
40 | 3.0 | 27494.98 | 25409.14 | 28382.98 | 21622.95 (21555.2, 21690.7) \\
40 | 4.0 | 13301.04 | 12846.72 | 13533.58 | 10726.94 (10696.9, 10757.0) \\
40 | 6.0 | 6015.13 | 6258.21 | 5894.10 | 5098.23 (5085.9, 5110.6) \\
40 | 8.0 | 3894.65 | 4241.90 | 3666.01 | 3320.64 (3313.1, 3328.1) \\

Table 6.5: Values of $SSANB(\gamma)$ for case 12 ($p_0 = 0.006$, $\gamma_1 = 1.25$, $m_0 = 400$), case 13 ($p_0 = 0.006$, $\gamma_1 = 2.25$, $m_0 = 350$), case 28 ($p_0 = 0.0009$, $\gamma_1 = 3.5$, $m_0 = 850$), and case 40 ($p_0 = 0.0002$, $\gamma_1 = 5.75$, $m_0 = 350$) for various hypothetical shifts of size $\gamma$. The 95% simultaneous confidence interval is $SSANB_b(\gamma) \pm 3.63 \times SE$ for case 40.

### 6.3 Comparison of the methods using the area under the $SSANB$ curves

In this section we use the area of under the log($SSANB(\gamma)$) curves to further summarize and compare the performance of the surveillance methods across a range of shifts. The motivation for this metric was discussed in more detail in Section 4.11. An approximation of $\int_{7.25}^{7.75} \log(SSANB(\gamma)) d\gamma$ was calculated using...
Simpson’s rule [40] for the four methods for each of the 40 cases. The results are shown in Table 6.6, where lower values indicate superior SSANB performance across the range of shifts. The upper limit of the integral approximation was 7.75 instead of 8 because Simpson’s rule requires that the domain over which the integral is calculated be partitioned into an even number of equally spaced intervals.

<table>
<thead>
<tr>
<th>Case</th>
<th>Sets</th>
<th>SHDA</th>
<th>CUSCORE</th>
<th>CUSUM</th>
<th>Case</th>
<th>Sets</th>
<th>SHDA</th>
<th>CUSCORE</th>
<th>CUSUM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>37.95</td>
<td>37.51</td>
<td>35.96</td>
<td>35.12</td>
<td>21</td>
<td>44.77</td>
<td>44.86</td>
<td>44.35</td>
<td>43.36</td>
</tr>
<tr>
<td>2</td>
<td>38.90</td>
<td>38.90</td>
<td>38.04</td>
<td>37.05</td>
<td>22</td>
<td>41.54</td>
<td>41.84</td>
<td>40.88</td>
<td>40.31</td>
</tr>
<tr>
<td>3</td>
<td>35.33</td>
<td>35.53</td>
<td>35.02</td>
<td>34.28</td>
<td>23</td>
<td>43.47</td>
<td>43.73</td>
<td>43.11</td>
<td>42.50</td>
</tr>
<tr>
<td>4</td>
<td>36.72</td>
<td>36.90</td>
<td>36.64</td>
<td>35.67</td>
<td>24</td>
<td>46.92</td>
<td>47.12</td>
<td>46.44</td>
<td>45.73</td>
</tr>
<tr>
<td>5</td>
<td>34.77</td>
<td>35.45</td>
<td>34.27</td>
<td>33.81</td>
<td>25</td>
<td>49.45</td>
<td>49.54</td>
<td>49.03</td>
<td>48.06</td>
</tr>
<tr>
<td>6</td>
<td>40.52</td>
<td>40.35</td>
<td>39.11</td>
<td>37.89</td>
<td>26</td>
<td>49.62</td>
<td>50.06</td>
<td>49.32</td>
<td>48.53</td>
</tr>
<tr>
<td>7</td>
<td>40.38</td>
<td>40.77</td>
<td>39.69</td>
<td>38.73</td>
<td>27</td>
<td>50.64</td>
<td>50.88</td>
<td>50.25</td>
<td>49.76</td>
</tr>
<tr>
<td>8</td>
<td>40.07</td>
<td>40.01</td>
<td>37.40</td>
<td>37.11</td>
<td>28</td>
<td>54.51</td>
<td>54.53</td>
<td>53.67</td>
<td>52.55</td>
</tr>
<tr>
<td>9</td>
<td>45.82</td>
<td>46.03</td>
<td>42.69</td>
<td>41.53</td>
<td>29</td>
<td>48.77</td>
<td>49.06</td>
<td>48.62</td>
<td>48.03</td>
</tr>
<tr>
<td>10</td>
<td>39.86</td>
<td>39.76</td>
<td>38.90</td>
<td>37.93</td>
<td>30</td>
<td>54.48</td>
<td>54.61</td>
<td>54.06</td>
<td>53.13</td>
</tr>
<tr>
<td>11</td>
<td>36.95</td>
<td>37.20</td>
<td>36.64</td>
<td>35.91</td>
<td>31</td>
<td>54.93</td>
<td>55.27</td>
<td>54.36</td>
<td>53.40</td>
</tr>
<tr>
<td>12</td>
<td>45.12</td>
<td>44.66</td>
<td>42.06</td>
<td>41.35</td>
<td>32</td>
<td>52.54</td>
<td>52.88</td>
<td>52.44</td>
<td>51.72</td>
</tr>
<tr>
<td>13</td>
<td>41.11</td>
<td>41.74</td>
<td>40.20</td>
<td>39.10</td>
<td>33</td>
<td>57.51</td>
<td>57.31</td>
<td>56.77</td>
<td>56.08</td>
</tr>
<tr>
<td>14</td>
<td>37.96</td>
<td>38.20</td>
<td>37.64</td>
<td>36.91</td>
<td>34</td>
<td>59.14</td>
<td>59.23</td>
<td>58.48</td>
<td>57.41*</td>
</tr>
<tr>
<td>15</td>
<td>37.84</td>
<td>38.52</td>
<td>37.55</td>
<td>36.88</td>
<td>35</td>
<td>57.53</td>
<td>57.60</td>
<td>57.02</td>
<td>56.29</td>
</tr>
<tr>
<td>16</td>
<td>42.24</td>
<td>42.57</td>
<td>41.38</td>
<td>40.39</td>
<td>36</td>
<td>56.46</td>
<td>56.66</td>
<td>55.48</td>
<td>54.93*</td>
</tr>
<tr>
<td>17</td>
<td>37.59</td>
<td>37.89</td>
<td>37.44</td>
<td>36.87</td>
<td>37</td>
<td>61.98</td>
<td>61.82</td>
<td>60.41</td>
<td>59.42*</td>
</tr>
<tr>
<td>18</td>
<td>37.59</td>
<td>37.89</td>
<td>37.44</td>
<td>36.99</td>
<td>38</td>
<td>64.02</td>
<td>64.10</td>
<td>63.38</td>
<td>62.29*</td>
</tr>
<tr>
<td>19</td>
<td>44.77</td>
<td>44.86</td>
<td>44.35</td>
<td>43.10</td>
<td>39</td>
<td>60.08</td>
<td>60.35</td>
<td>60.36</td>
<td>59.31*</td>
</tr>
<tr>
<td>20</td>
<td>45.05</td>
<td>45.15</td>
<td>44.64</td>
<td>43.47</td>
<td>40</td>
<td>62.84</td>
<td>62.76</td>
<td>62.87</td>
<td>61.65*</td>
</tr>
</tbody>
</table>

Table 6.6: Approximate areas under the log(SSANB) curves for the 40 cases ($\approx \int_{1.25}^{7.75} \log(\text{SSANB}(\gamma))d\gamma$). For the simulated cases indicated by the asterisk, the area was computed under the upper bound of the 95% simultaneous confidence interval ($\approx \int_{1.25}^{7.75} \log(\text{SSANB}_b(\gamma) + 3.63 \ SE)d\gamma$).

Table 6.7 summarizes the information in Table 6.6. In all 40 cases, the CUSUM had the smallest area under the log(SSANB) curve, further confirming the CUSUM as having the best SSANB performance across a range of shifts, regardless of the size of shift for which the CUSUM was designed to detect. Further examination of Table 6.7 reveals a pattern that is similar to the inequality in (6.1), where in the vast majority of the cases, the CUSCORE shows the second best performance, followed by the Sets method with the third best performance, and the SHDA method with the worst performance among the four methods.
Table 6.7: Summary of Table 6.6, indicating the number of cases for each method where $\int_{1.25}^{7.75} \log(SSANB(\gamma)) d\gamma$ was least among the four methods (“Best”), second least among the four methods (“2nd best”), etc.

<table>
<thead>
<tr>
<th>Method</th>
<th>Best</th>
<th>2nd best</th>
<th>3rd best</th>
<th>4th best</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sets</td>
<td>0</td>
<td>1</td>
<td>32</td>
<td>7</td>
</tr>
<tr>
<td>SHDA</td>
<td>0</td>
<td>1</td>
<td>8</td>
<td>31</td>
</tr>
<tr>
<td>CUSCORE</td>
<td>0</td>
<td>38</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>CUSUM</td>
<td>40</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

6.4 Effects of $p_0$, $\gamma_1$, and $m_0$ on chart performance

One of the reasons why so many different values of $p_0$, $\gamma_1$, and $m_0$ were considered was to assess the effect of these variables on the performance of the CUSUM chart relative to the other methods. Since the CUSCORE had the second best SSANB performance in 1069 of the 1160 comparisons of the four methods (92%) (and the best SSANB performance in 13 of the comparisons, see Table 6.4), and since it also had the second lowest area under the log(SSANB) curves in 38 of the 40 cases, we chose the CUSCORE as the method to compare against the CUSUM in order to investigate the influence of $p_0$, $\gamma_1$, and $m_0$ on SSANB performance. We used multiple regression with predictors $p_0$, $\gamma_1$, and $m_0$ to model the following response variables which were calculated for each of the 40 cases:

\[
y_1 = \log(SSANB_c(\gamma_{1a})) - \log(SSANB_b(\gamma_{1a}))
\]

\[
y_2 = \int_{1.25}^{7.75} \log\left(\frac{SSANB_c(\gamma)}{SSANB_b(\gamma)}\right) d\gamma.
\]

$y_1$ is the difference in the log(SSANB) between the CUSCORE and the CUSUM, evaluated at the shift for which the chart was designed. $y_2$ is the difference in the areas underneath the log(SSANB) curves between the CUSCORE and the CUSUM. The larger the values of $y_1$ and $y_2$, the better the CUSUM is performing with respect to the CUSCORE. For both $y_1$ and $y_2$, the initial regression model that we considered (with subscripts for individual data points omitted) is given below:

\[
y = \beta_0 + \beta_1 p_0 + \beta_2 \gamma_1 + \beta_3 m_0 + \beta_4 \gamma_{1a}^2 + \beta_5 m_{0}^2 + \beta_6 p_0 \gamma_{1a} + \beta_7 p_0 m_0 + \beta_8 \gamma_{1a} m_0 + \beta_9 p_0 \gamma_{1a} m_0 + \epsilon.
\]

To achieve parsimonious models, coefficients in the initial fitted model that were not significant at the 5% level were successively removed, beginning with the higher order terms first. None of the interactions for either response were significant. The final models are given by equation (6.2) and their estimated coefficients
and corresponding significance tests are shown in Table 6.8.

\[
\begin{align*}
    y_1 &= \beta_0 + \beta_1 p_0 + \beta_2 \gamma_1 a + \beta_3 m_0 + \beta_4 m_0^2 + \epsilon \\
    y_2 &= \beta_0 + \beta_1 p_0 + \beta_2 \gamma_1 a + \beta_3 m_0 + \beta_4 \gamma_1 a^2 + \beta_5 m_0^2 + \epsilon
\end{align*}
\]  

(6.2)

Analysis of the residuals did not reveal any striking outliers, nor did it show a lack of homogeneity of

| Response | Predictor | $\hat{\beta}$ | SE(\hat{\beta}) | t value | 2P(>|t|) |
|----------|-----------|---------------|-----------------|---------|----------|
| $y_1$    | Intercept | 0.19186766    | 0.01168731      | 16.417  | 0.0000   |
|          | $p_0$     | -0.49435683   | 0.92085589      | -0.537  | 0.5948   |
|          | $\gamma_1 a$ | -0.02329208  | 0.00196319      | -11.864 | 0.0000   |
|          | $m_0$     | 0.00034190    | 0.00004081      | 8.377   | 0.0000   |
|          | $m_0^2$   | -0.00000027   | 0.00000005      | -5.930  | 0.0000   |
| $y_2$    | Intercept | 0.09230039    | 0.18104876      | 0.510   | 0.6135   |
|          | $p_0$     | 2.94248603    | 8.29234002      | 0.355   | 0.7249   |
|          | $\gamma_1 a$ | 0.24695958   | 0.09308050      | 2.653   | 0.0120   |
|          | $m_0$     | 0.00183229    | 0.00037204      | 4.925   | 0.0000   |
|          | $\gamma_1 a^2$ | -0.03317481 | 0.01249843      | -2.654  | 0.0120   |
|          | $m_0^2$   | -0.00000139   | 0.00000041      | -3.356  | 0.0020   |

Table 6.8: Fitted regression models to assess the effect of $p_0$, $\gamma_1 a$, and $m_0$ on the responses $y_1$ and $y_2$. variance nor a serious departure from normality in either model. The $R^2$ values for the fitted models were 0.89 and 0.67 for $y_1$ and $y_2$, respectively.

![Graph](image1)

Figure 6.4: Plots of $y_1$ versus $p_0$, $\gamma_1 a$, and $m_0$ with fitted values superimposed. In each plot, the fitted regression line was calculated by allowing the predictor variable of interest to vary while the other two predictor variables were fixed at their respective means.

Figures 6.4 and 6.5 show the plots of the responses $y_1$ and $y_2$ versus the predictors $p_0$, $\gamma_1 a$, and $m_0$ with the fits of the regression models superimposed. For both $y_1$ and $y_2$, the baseline incidence rate $p_0$
Figure 6.5: Plots of $y_2$ versus $p_0$, $\gamma_1 a$, and $m_0$ with fitted values superimposed. In each plot, the fitted regression line was calculated by allowing the predictor variable of interest to vary while the other two predictor variables were fixed at their respective means.

does not affect the difference in performance between the CUSCORE and the CUSUM. Both responses are nicely modeled as a concave quadratic function of the target in-control average number of malformations until signal, $m_0$, with maximal responses predicted at 633 and 659 malformations until signal for $y_1$ and $y_2$ respectively. This suggests that the improvement in performance of the CUSUM over the CUSCORE increases as $m_0$ increases from 50 to 600, and shortly thereafter the improvement moderately declines. In Figure 6.4, the middle plot indicates that the improvement in the performance of the CUSUM over the CUSCORE at detecting the shift for which the chart was designed is highest for small shifts, with the difference in performance decreasing linearly as $\gamma_1$ increases. The middle plot in Figure 6.5 demonstrates a quadratic relationship between the difference in the area underneath the log(SSANB) curves and the shift for which the chart was designed ($\gamma_1 a$). The CUSUM's improvement in performance over the CUSCORE is predicted to be highest when designing charts to detect a 3.7 fold increase in the baseline incidence rate.
Chapter 7

Discussion

7.1 Further considerations in assessing chart performance

In order to fairly compare two or more surveillance methods, the criteria for choosing the design parameters, the metric used to compare the methods, and other features of the method (like a head-start) must be carefully considered. We discuss the choice of design parameters in Section 7.2 and focus our discussion in this section on the performance metrics and head-start features of the four methods.

Most of the research surrounding the Sets, SHDA, and CUSCORE methods has been based on the assumption that the data arise from geometric random variables (number of births between malformations) or as a Poisson process (amount of time elapsed between incidents) [7, 8, 9, 16, 17, 18, 19, 21]. This view of the data, however, can lead to the assumption that any increase in the incidence rate occurs immediately following the observation of a malformation, which is unrealistic if the data are viewed as a sequence of Bernoulli variables. Furthermore, the head-start of the Sets and CUSCORE methods only becomes apparent when the data are modeled as Bernoulli variables. The head-start of the SHDA method however, is present regardless of whether the data are viewed as geometric or Poisson variates or as Bernoulli trials, since the initial state in which monitoring begins is not the furthest away from the absorbing (signaling) state.

7.1.1 Measuring performance using the initial state ANB

For the Sets, SHDA, and CUSCORE methods, unless we suspect that the epidemic is already underway at the time that monitoring begins, neither the initial state average number of malformations until signal \( ANM(p) \) nor the initial state average number of births until signal \( ANB(p) \) are realistic measures of
chart performance because of the head-start features of these three methods. We note that using $ANB(p)$ to compare the four methods requires the following assumptions:

1. The shift in the incidence rate has taken place before monitoring begins.

2. For the Sets, SHDA, and CUSCORE methods, the first set size is calculated as if a malformed case were observed immediately prior to the onset of monitoring.

3. For the SHDA method, the chart statistics are started as if a flag were raised on the malformed case that is assumed to occur just before monitoring begins.

Among the references that are cited herein, only Wolter [8] addresses the first assumption by comparing the Sets and the CUSCORE methods using the steady-state average number of malformations until signal, $SSANM$ (where the steady-state distribution is not conditional on not having observed a false alarm). However, using the $SSANM$ as the performance metric is based on the assumption that the shift occurs immediately following the observation of a malformed case, a rare event that occurs with probability $p_0$. Assumptions 2 and 3, which are consequences of the head-start features implicit in the sets based methods, were never explicitly mentioned by the authors of the cited references. This was a major weakness in the evaluation of the SHDA and Sets methods by Sitter et al. [9]. Note that assumption 3 adds an additional head-start to the SHDA method by starting the chart in a state that is not furthest from signaling, whereas the Sets and CUSCORE methods initialize with $S_0 = 0$ and $C_0 = 0$, respectively, which are the states that are furthest from signal. We further add that assumptions 2 and 3 are not applicable to the CUSUM because it does not require grouping the Bernoulli observations into sets before incorporating information about the process into the chart.

Since the Bernoulli CUSUM that we have considered does not have a head-start feature, if we were to use the $ANB(p_1)$ as the only metric for our comparisons, we would be led to make naive and even erroneous conclusions. Table 7.1 gives the initial state average number of births until signal evaluated at the shift for which the chart was designed, $ANB(\gamma_1a p_0)$. The Ratio column in Table 7.1 gives the $ANB_b(\gamma_1a p_0)$ of the CUSUM divided by the minimum $ANB(\gamma_1a p_0)$ of all four methods in the corresponding case. If we calculate the average of the ratio across the forty cases, we find that the $ANB$ performance is only about 1.4 times worse on average than the best performing method, which is noteworthy considering the absence of a head-start feature in the CUSUM. The results of Table 7.1 are summarized in Table 7.2. Owing to the head-start of the sets based methods, the CUSUM has the best $ANB$ performance when the actual shift matches the shift for which chart was designed in only 4 of the 40 cases. Due to its double-layer head-start
<table>
<thead>
<tr>
<th>Case</th>
<th>ANB(γa0)</th>
<th>Ratio</th>
<th>95% Simul. C.I.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sets</td>
<td>SHDA</td>
<td>CUSCORE</td>
</tr>
<tr>
<td>1</td>
<td>2355.45</td>
<td>1810.68</td>
<td>2211.47</td>
</tr>
<tr>
<td>2</td>
<td>299.74</td>
<td>175.53</td>
<td>279.66</td>
</tr>
<tr>
<td>3</td>
<td>207.85</td>
<td>131.19</td>
<td>201.38</td>
</tr>
<tr>
<td>4</td>
<td>161.80</td>
<td>93.87</td>
<td>158.16</td>
</tr>
<tr>
<td>5</td>
<td>724.56</td>
<td>509.25</td>
<td>720.16</td>
</tr>
<tr>
<td>6</td>
<td>1374.09</td>
<td>914.78</td>
<td>1240.16</td>
</tr>
<tr>
<td>7</td>
<td>265.63</td>
<td>166.04</td>
<td>249.88</td>
</tr>
<tr>
<td>8</td>
<td>5376.46</td>
<td>4533.77</td>
<td>5163.31</td>
</tr>
<tr>
<td>9</td>
<td>19804.75</td>
<td>16625.90</td>
<td>15789.47</td>
</tr>
<tr>
<td>10</td>
<td>777.48</td>
<td>503.66</td>
<td>731.67</td>
</tr>
<tr>
<td>11</td>
<td>276.87</td>
<td>178.66</td>
<td>268.79</td>
</tr>
<tr>
<td>12</td>
<td>1574.87</td>
<td>1021.15</td>
<td>1454.50</td>
</tr>
<tr>
<td>13</td>
<td>317.10</td>
<td>202.80</td>
<td>306.69</td>
</tr>
<tr>
<td>14</td>
<td>1153.79</td>
<td>810.82</td>
<td>1139.51</td>
</tr>
<tr>
<td>15</td>
<td>1487.62</td>
<td>957.90</td>
<td>1367.84</td>
</tr>
<tr>
<td>16</td>
<td>210.31</td>
<td>125.63</td>
<td>208.11</td>
</tr>
<tr>
<td>17</td>
<td>451.27</td>
<td>266.85</td>
<td>428.94</td>
</tr>
<tr>
<td>18</td>
<td>382.56</td>
<td>220.92</td>
<td>363.57</td>
</tr>
<tr>
<td>19</td>
<td>351.00</td>
<td>198.75</td>
<td>332.52</td>
</tr>
<tr>
<td>20</td>
<td>2678.73</td>
<td>2013.94</td>
<td>2637.29</td>
</tr>
<tr>
<td>21</td>
<td>341.91</td>
<td>200.49</td>
<td>337.71</td>
</tr>
<tr>
<td>22</td>
<td>515.28</td>
<td>304.51</td>
<td>501.77</td>
</tr>
<tr>
<td>23</td>
<td>657.72</td>
<td>369.64</td>
<td>622.51</td>
</tr>
<tr>
<td>24</td>
<td>2536.97</td>
<td>1553.22</td>
<td>2498.88</td>
</tr>
<tr>
<td>25</td>
<td>885.06</td>
<td>505.34</td>
<td>878.17</td>
</tr>
<tr>
<td>26</td>
<td>4477.60</td>
<td>2817.41</td>
<td>4129.74</td>
</tr>
<tr>
<td>27</td>
<td>1163.37</td>
<td>694.73</td>
<td>1151.54</td>
</tr>
<tr>
<td>28</td>
<td>2711.14</td>
<td>1582.11</td>
<td>2589.65</td>
</tr>
<tr>
<td>29</td>
<td>6823.10</td>
<td>4199.33</td>
<td>6503.48</td>
</tr>
<tr>
<td>30</td>
<td>3385.77</td>
<td>2092.16</td>
<td>3376.90</td>
</tr>
<tr>
<td>31</td>
<td>2502.71</td>
<td>1415.83</td>
<td>2425.21</td>
</tr>
<tr>
<td>32</td>
<td>7927.35</td>
<td>4698.32</td>
<td>7402.31</td>
</tr>
<tr>
<td>33</td>
<td>3900.19</td>
<td>2254.47</td>
<td>3820.46</td>
</tr>
<tr>
<td>34</td>
<td>29000.63</td>
<td>21576.95</td>
<td>27948.14</td>
</tr>
<tr>
<td>35</td>
<td>38014.80</td>
<td>25585.25</td>
<td>34071.03</td>
</tr>
<tr>
<td>36</td>
<td>14561.26</td>
<td>8546.94</td>
<td>13561.03</td>
</tr>
<tr>
<td>37</td>
<td>5042.82</td>
<td>2900.11</td>
<td>5019.56</td>
</tr>
<tr>
<td>38</td>
<td>5921.19</td>
<td>3372.66</td>
<td>5770.14</td>
</tr>
</tbody>
</table>

Table 7.1: Values of $ANB(\gamma a0)$ of the four methods evaluated for each of the 40 cases. For the six simulated cases, the 95% simultaneous confidence interval is given by $ANB(\gamma a0) \pm 3.63 \times SE$. The Ratio column is the CUSUM $ANB(\gamma a0)$ divided by the lowest $ANB(\gamma a0)$ among the four methods for the given case.
<table>
<thead>
<tr>
<th>Method</th>
<th>Best</th>
<th>2nd best</th>
<th>3rd best</th>
<th>4th best</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sets</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>32</td>
</tr>
<tr>
<td>SHDA</td>
<td>36</td>
<td>2</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>CUSCORE</td>
<td>0</td>
<td>10</td>
<td>30</td>
<td>0</td>
</tr>
<tr>
<td>CUSUM</td>
<td>4</td>
<td>28</td>
<td>0</td>
<td>8</td>
</tr>
</tbody>
</table>

Table 7.2: Summary of Table 7.1, indicating the number of cases for each method where the $ANB(\gamma_1a)_{p_0}$ was least among the four methods ("Best"), second least among the four methods ("2nd best"), etc.

feature, it is not surprising that the SHDA method has the best $ANB(\gamma_1a)_{p_0}$ performance among the four methods in 36 of the 40 cases. Further examination of Table 7.2 shows the CUSUM having the second best performance in 28 of the 40 cases, the CUSCORE having third best performance in 30 cases, and the Sets method most often having the worst $ANB(\gamma_1a)_{p_0}$ performance.

Obviously the conclusions one would draw by evaluating the methods solely on the basis of $ANB(\gamma_1a)_{p_0}$ performance would be radically different than those presented in Chapter 6. We feel the assumptions imposed by using the $ANB(p_1)$ as a performance measure are both restrictive and unrealistic.

7.1.2 Measuring performance using the $SSANB$

Consider instead the assumptions implicit in using $SSANB$ as a comparative measure of performance:

1. The chart has not signaled prior to the shift (i.e. no false alarms).

2. The shift occurs after the chart has converged to the steady-state distribution.

The first assumption removes the need to account for false alarms that may occur prior to the shift and permits us to focus only on the average run length until the first alarm. If we were to use the unconditional steady-state distribution and map the absorbing states back to the initial states of the chart (thereby resetting the chart after a false alarm), we would be assuming that, in practice, we could clearly distinguish between false and real alarms. Determining the validity of an alarm may be possible in an industrial process control setting, where the process may be temporarily shut down for investigation. In this situation, restarting the chart in the initial state is certainly legitimate. However, in an epidemiological setting, one does not simply stop the “process,” nor is it as readily apparent whether or not an alarm is false (Chen and others have discussed and proposed methods for assessing the nature of an alarm [7, 14, 41]). For purposes of comparing surveillance methods, assuming that the chart has not signaled prior to the shift in the incidence rate permits us to focus simply on the average run length since the shift until the first alarm.
The second assumption for using the SSANB as a performance metric is that the steady-state distribution has been achieved prior to the occurrence of the shift. To assess the impact of this assumption, we investigated the convergence of \( WANB(\tau) \) (the weighted average of the average number of births from a shift at time \( \tau \) until signal) to the SSANB. Refer to Section 4.10 for a discussion and derivation of \( WANB(\tau) \). Figure 7.1 graphically demonstrates the \( WANB \) convergence for each of the four methods for case 11. Recall that for the three sets based methods, we let the value \( i = -1 \) indicate that the shift has taken place before monitoring begins, \( i = 0 \) indicate that the shift takes place between the start of monitoring and the observance of the first malformation and \( i \geq 1 \) indicate that the shift has taken place between malformation \( i \) and \( i + 1 \). For the CUSUM, the index \( j \) indicates the shift has taken place between births \( j \) and \( j + 1 \). In Figure 7.1, \( WANB(\tau) \) and the SSANB were evaluated at \( \gamma_{1a} = 3.70731 \). Convergence was declared to have occurred when the ratio \( \frac{WANB(\tau)(\gamma_{1a})}{SSANB(\gamma_{1a})} \) fell and remained within the interval \((0.999, 1.001)\). This interval is depicted by the horizontal dashed lines in the plots. The top horizontal axes measure the rate of convergence as a fraction of the in-control average number of births until signal, \( ANB(p_0) \). To calculate this fraction for the sets based methods, the number of births before the signal was approximated as \( 1/p_0 \) births per incident.

One of the most notable features of the four plots in Figure 7.1 is that \( WANB \) converges very rapidly (with respect to \( ANB(p_0) \)) for each of the four methods—all of them converge within 7% of the \( ANB(p_0) \). The figure also allows us to visualize the nature of the head-start for the sets based methods. Note that for the Sets and CUSCORE methods, the effect of the head-start diminishes very quickly. However, the effect of the head-start for the SHDA method lasts much longer—the length of the effect depending on the design parameters \( n_t \) and \( b_t \). The head-start for the SHDA method effectively vanishes once (and if) the first flag is raised without signaling. The shape of the plots in Figure 7.1 is typical for the other 39 cases that were considered.

Table 7.3 shows the number of births as a fraction of the \( ANB(p_0) \) that are required for convergence as previously defined. The convergence rate was calculated for each of the 40 cases for the sets based methods, and for the 29 cases of the CUSUM that were computationally feasible. For the CUSUM, calculating \( WANB_b^{(j)}(\gamma) \) for each birth can be computationally expensive, especially if the number of states is large. To reduce the number of matrix multiplications, a matrix \( Q_b^J \), representing the observation of \( J \) births was used to calculate \( WANB_b^{(j)}(\gamma) \) for \( j = \{0, 1, J + 1, 2J + 1, 3J + 1, \ldots\} \). This pattern is visible in the points of the CUSUM plot in Figure 7.1, where \( J = 64 \). As a general rule, \( J \) was chosen as the value in the sequence \( \{2^k : k = 1, 2, 3, \ldots\} \) that was near \( 1/p_0 \), the average number of births per malformation before
Figure 7.1: Convergence of WANB to SSANB for case 11 ($p_0 = 0.007$, $\gamma_1 = 3.75$, $m_0 = 100$). For the top horizontal axis of the Sets, SHDA, and CUSCORE methods, the number of births was approximated as $1/p_0$ births per incident. The horizontal dashed lines denote the bands inside which convergence was declared to have occurred. The vertical dotted lines mark the step when convergence occurred.

the shift occurs. This implies that the estimated times at which $WANB_b^{(j)}(\gamma_{1a})$ converges to $SSANB_b(\gamma_{1a})$ that are shown in Table 7.3 are upper bounds for the actual convergence time. Note at the bottom of the table that the mean and median number of births as a fraction of the $ANB(p_0)$ that were required for convergence is very small for all four of the methods, with the CUSCORE typically taking the longest time to converge. Figure 7.2 shows the distribution of the convergence rates for each of the methods. Given
the rapid convergence of $WANB(\tau)$ to $SSANB$, the assumption that the shift occurs after the steady-state distribution has been achieved is quite reasonable, thereby confirming the $SSANB$ as a more realistic measure of chart performance than $ANB(p_1)$.

![Sets, SHDA, CUSCORE, CUSUM Histograms](image)

Figure 7.2: Histograms of the number of births as a fraction of $ANB(p_0)$ at which convergence to $SSANB(\gamma_{1a})$ occurs.

### 7.1.3 Measuring performance for an arbitrary shift time

If one is willing to specify a distribution for the shift time $\tau$, the assumption that the shift occurs after convergence to the steady-state distribution is no longer necessary. One possible distribution for $\tau$ would be the mixture of a point mass corresponding to the shift occurring prior to the onset of monitoring with a uniform kernel beginning at the onset of monitoring and extending to a time in the distant future (perhaps near $ANB(p_0)$). Another possibility would be a mixture of the aforementioned point mass with an exponential kernel that decays slowly so that shifts in the very distant future are less likely than those that
are more proximal. After specifying the distribution of \( \tau \), one could then compare the methods (for a given shift \( \gamma \)) using

\[
M_1 = \int W \, ANB^{(\tau)}(\gamma) dT
\]  

(7.1)

where \( T \) is the probability measure associated with the distribution of \( \tau \). In order to calculate \( M_1 \), \( T \) would have to be suitably discretized so that the distribution of \( \tau \) would consist of a collection of point masses across the desired range of \( \tau \). However, unless the distribution of \( \tau \) were to place a relatively high amount of probability on the shift occurring either before or during the initial stages of monitoring, the quantity \( M_1 \) is likely to be close to \( SSANB(\gamma) \) since convergence to the steady-state distribution is rapid in most instances.

7.2 Criteria for choosing the design parameters

For the purposes of comparing two or more surveillance methods, there are a number of reasonable ways in which the design parameters of may be chosen. For simplicity in the discussion that follows, each of the approaches we present assumes that the values of the design parameters are chosen to minimize some function of the out-of-control \( ANB \) subject to the in-control \( ANB \) being at least as large as a specified constant, i.e. \( ANB(p_0) \geq b_0 \).

One approach that is commonly advocated is to minimize the out-of-control initial state average run length \( (ANB(p_1)) \) for a given step shift \( \gamma_1 \). The comparison of chart performance could then be based upon \( ANB(p_1) \) calculations. This is certainly a viable strategy if the methods under consideration do not have head-starts. However, designing and comparing methods that do have head-starts in this way can lead to misleading conclusions about chart performance.

Another approach is the “minimax” optimality criteria advocated by Lorden [37] and Moustakides [38], where \( W \), the out-of-control average run length that results when the specified shift takes place at the worst possible time and when the chart is in the least advantageous state, is minimized. Since the CUSUM is known to be optimal in this sense, we designed the sets based methods using this criteria in order to make comparisons with the CUSUM as equitable as possible.

Another possible approach would be to minimize the steady-state average run length \( (SSANB(\gamma_1)) \) for the shift \( \gamma_1 \). This approach was especially appealing to us, since we used \( SSANB \) as the primary metric.
to compare the four methods. Designing the Sets, SHDA, and CUSCORE methods using this criteria is certainly feasible, since the number of Markov chain states they contain is relatively small. However, finding the CUSUM reference value, $\delta$, that minimizes the $SSANB$ would be extremely computationally intensive at best and unfeasible in many instances. This is because each of the potentially many iterations required to find the optimal value of $\delta$ would involve inverting the very large matrix $I - Q_b$. We recognize that using the $SSANB$ to both design the chart and measure performance would be ideal. However, relying on the theoretical result given by Moustakides made it possible to quickly design thousands of CUSUM charts over a range of $p_0, \gamma_1$ and $m_0$ values in order to identify the specific cases where the $ANB(p_0)$ was most closely matched for the four methods.

Since the actual shift is unknown, another approach would be to search for design parameters which optimize performance over a range of possible shifts. Radaelli [26] discusses this idea, using the Sets method as an example. A potential drawback of this approach is the necessity to specify the distribution of the potential shifts. One possibility would be a uniform distribution on a specified interval $(a, b)$. Another would be a distribution that is constant over the interval $(a, b)$ and then decays according to the right half of the Gaussian kernel over the interval $[b, \infty)$. If small shifts are assumed to be more likely than larger shifts, an exponential (or truncated exponential) distribution could also be used. Once the distribution of $\gamma$ is specified, the design parameters would then be chosen to minimize

$$M_2 = \int SSANB(\gamma)d\Gamma,$$

where $\Gamma$ is the measure for the distribution of $\gamma$.

The final approach that we consider would be to choose the design parameters to be optimal not only over a range of shift sizes but also across a range of shift times. This would be pertinent for evaluating charts that have head-start features, since the improved performance associated with the head-start during the initial period of monitoring would not be excluded from the parameter selection process, as it is for the previous three approaches. As with the quantities $M_1$ and $M_2$ (see equations (7.1) and (7.2)), a distribution would have to be assumed both for the shift time $\tau$ and the shift size $\gamma$ with corresponding probability measures $T$ and $\Gamma$. The design parameters would then be chosen to minimize

$$M_3 = \int \int WANB(\tau, \gamma)dTd\Gamma,$$
where $WANB^{(r)}(\gamma)$ is discussed in Section 4.10. Unfortunately, just like the $SSANB$, using $M_2$ or $M_3$ as the basis for parameter selection has the potential of being very computationally intensive or even unfeasible, particularly for the CUSUM chart.

7.3 Regarding potential bias against the sets based methods

7.3.1 Effect of the optimality criteria used to select design parameters

It could be argued that the comparisons presented in Chapter 6 naturally favored the CUSUM because the sets based methods were designed to be optimal in the same way in which the CUSUM is optimal. Perhaps a more equitable comparison would involve choosing the design parameters that minimize the initial state $ANB(p_1)$ (rather than minimizing $W$, the worst case $ANB(p_1)$) and then evaluating the charts using the $SSANB$. For the Sets and the CUSCORE methods, this argument is not applicable because the values of the design parameters that minimize $ANB(p_1)$ for these two methods are the same as those that minimize $W$. To see this, we refer the reader back to equations (5.2) and (5.4) and the statements that immediately follow these equations. For the CUSUM chart, since the initial state $ANB_0(p_1)$ is equal to the worst case $ANB_0(p_1)$, the same value of $\delta$ minimizes both quantities. However, for the SHDA method, the design parameters that minimize $ANB(t)(p_1)$ are different from those that minimize $W_t$.

In a smaller study we performed previously, we compared the $SSANB$ performance of the four methods when the design parameters for each chart were chosen to minimize the initial state $ANB(p_1)$. This study consisted of 23 different combinations (cases) of the baseline incidence rate $p_0$, the target in-control average number of births until signal $b_0$, and the shift size for which the chart was designed to be optimal, $\gamma_1$. The values of $p_0$ were 0.01, 0.005, 0.001, 0.0005, and 0.0001. The target in-control $ANB$ ranged from $b_0 = 10,000$ to $b_0 = 500,000$, and the shift sizes for which the charts were designed to be optimal (with respect to minimizing the $ANB(\gamma_1 p_0)$) were $\gamma_1 = 2, 4,$ and $6$. The control limits were chosen so that $ANB(p_0)$ was as close to $b_0$ as possible while still preserving the inequality $ANB(p_0) \geq b_0$. For each combination, $SSANB(\gamma)$ was evaluated at $\gamma = \{1.25, 1.5, \ldots, 7.75, 8\}$.

The results of this small study were very similar to those of the larger study presented in Section 6.2. The $SSANB$ performance of the CUSUM at the shift for which the charts were designed was best among the four methods for all 23 cases. That is, $SSANB_0(\gamma_1 a) < \min[SSANB_a(\gamma_1 a), SSANB_t(\gamma_1 a), SSANB_c(\gamma_1 a)]$. In addition, the inequality in (6.1) held for 21 of the 23 cases. Among the 644 instances in which $SSANB(\gamma)$ was calculated (23 cases $\times$ 28 values of $\gamma = 644$), the CUSUM had the best $SSANB$ performance in 631 of
those instances. The 13 instances in which the CUSUM was not best are shown in Table 7.4. Note that the CUSCORE and SHDA methods had better $SSANB$ performance when the shift for which the chart was designed is large ($\gamma_1 = 6$) and the actual shift size is small ($\gamma \leq 2$).

In summary, when the design parameters are chosen to minimize the initial state $ANB(p_1)$, the $SSANB$ performance of the Bernoulli CUSUM is almost uniformly better than the other methods for the cases we considered. The preservation of inequality (6.1) in almost all of the 23 cases suggests that the CUSUM has the best $SSANB$ performance at the shift for which the chart was designed, followed by the CUSCORE, then the Sets method, and last of all, the SHDA method. The results in Table 7.4 suggests that choosing the design parameters of the SHDA method to to minimize $ANB_t(p_1)$ may improve the $SHDA$ performance slightly, but certainly not enough to appreciably alter the results given in Chapter 6.

7.3.2 Regarding the head-start feature of the sets based methods

Just as the comparison of methods by using the initial state $ARL$ favors charts with head-starts, the comparison of methods using the steady-state $ARL$ can favor charts that do not have head-starts. Due to the head-starts of the sets based methods, the value of the threshold $t$ that satisfies $ANM(p_0) \geq m_0$ is lower than it otherwise would be if they did not have head-starts. Note that a lower value of $t$ translates into larger values of $ANB(p)$ and, most importantly, larger values of $SSANB$. Thus, in order to achieve the required value of the target in-control $ARL$, any chart with a head-start feature pays a price if the shift occurs after the head-start has worn off because it is not able to signal as quickly as an otherwise equivalent chart without a head-start. This is why the Bernoulli CUSUM does have an advantage over the sets based methods when the comparisons are based on the $SSANB$. Lucas and Crosier [42] examined the effect of adding a head-start to a CUSUM chart based on normally distributed data. They find that the price incurred by the head-start is small relative to the performance gains in detecting a shift that occurs prior to monitoring. However, we point out that when the shift occurs after the steady-state distribution has been reached, a chart with a head-start does not detect shifts as quickly as the same chart (with the same in-control $ARL$) that does not have a head-start.

At this point it is important to distinguish between the head-start of the Sets and CUSCORE methods and the head-start of the SHDA method. The Sets and CUSCORE methods both begin monitoring in the Markov chain state that is furthest away from signal (recall that $S_0 = 0$ and $C_0 = 0$). Their head-starts only exist when compared to the Bernoulli CUSUM, because the Sets and CUSCORE methods assume that a malformation was observed just prior to the onset of monitoring, and the Bernoulli CUSUM does not.
To make the SSANB comparisons of the Sets and CUSCORE methods versus the Bernoulli CUSUM more equivalent, we could initialize the Bernoulli CUSUM at $B_0 = m - 1$, which is tantamount to assuming that a malformation was observed immediately prior to the beginning of monitoring with the CUSUM.

The SHDA method also assumes that a malformation was observed just prior to monitoring. However, it differs from the other two sets based methods because it additionally assumes that a flag was raised just prior to the onset of monitoring. This results in the chart beginning in a Markov chain state that is actually closer to signaling than some of the other states in the chart. In this sense, the SHDA method has a “true” head-start, irrespective of other methods to which it might be compared. When using the SSANB to measure chart performance, the penalty incurred by the head-start of the SHDA method can effectively be removed if the threshold $t_t$ is chosen so that the in-control average number of births until signal when the chart begins in the state that is furthest from signal, $ANB_t(p_0)_{\text{furthest}}$, is at least as large as $b_0$. This is possible because the steady-state distribution does not depend on the state where the chart begins. Note that $ANB_t(p_0)_{\text{furthest}}$ can be found by slightly modifying equation (5.3) to give

$$ANB_t(p_0)_{\text{furthest}} = \frac{1}{p_0} \left( 1 - \frac{\theta_t^{\text{ins}}}{\theta_t^{\text{ins}}(1 - \theta_t) \left( 1 + \frac{1}{\Psi} \right)} \right)$$

(7.4)

where $\Psi$ is given by equation (4.8) and $\theta_t$ is calculated under incidence rate $p_0$. Therefore, if we choose the value of $t_t$ that satisfies $ANB_t(p_0)_{\text{furthest}} \geq b_0$, the resulting SSANB will be equivalent to the SHDA method with no head-start.

We use case 26 from the comparisons presented in Chapter 6 to illustrate the impact of the “price” incurred by the head-starts of the sets based methods. We will do this by attempting to adjust the CUSUM and the SHDA methods so that the penalty induced by the head-starts of the sets based methods is effectively removed. To adjust for the head-start that results from assuming the birth prior to the onset of monitoring was diagnosed with a malformation, the initial value of the CUSUM was set to $B_0 = m - 1 = 523$. The parameter $m$ was kept constant at 524, and $w$, the number of states, was increased from 1,926 to 1,930 to achieve $N_{523} = 100,198.3$, the in-control average number of births until signal when the chart is started in state 523. (Equivalently, we could say that the control limit $h$ was increased from 3.67557 to 3.68321). To adjust for the head-start of the SHDA method that results from the initial state being closer to signal than some of the other states, we increased the threshold $t_t$ from 714 to 740 so that $ANB_t(0.001)_{\text{furthest}} = 100,283.5$. Since compensation for the head-starts of the Sets and CUSCORE methods was accomplished by adjusting the CUSUM chart, we make no adjustments to the Sets
or CUSCORE methods. All the remaining design parameters not mentioned above were held fixed for the four methods. We compare the results of the adjustment using \( SSANB \) evaluated at \( \gamma_1 = 3.25 \), which is very close to the shift size for which the charts were designed to be optimal for case 26. The results are shown in Table 7.5.

The absolute difference between the adjusted and unadjusted \( SSANB \) values for the CUSUM shown in Table 7.5 can be thought of as the “price” incurred by the head-start which results from assuming that a malformation was observed just before monitoring. The absolute difference between the adjusted and unadjusted \( SSANB \) values for the SHDA method can be thought of as the “price” incurred by head-start of the SHDA method that results from starting the chart in a state that is not furthest from signal. When we remove the head-start from the SHDA method, its steady-state performance appears to be more comparable to the Sets and CUSCORE methods. However, the Bernoulli CUSUM continues to have the best \( SSANB \) performance among the four methods. It appears that the overall penalty incurred by the Sets and CUSCORE methods because of their head-starts when we assume the shift has occurred after convergence to the steady-state distribution is minimal. The penalty incurred by the SHDA method is more substantial—-but not enough to alter the conclusion that the \( SSANB \) performance of the CUSUM is superior to the other methods.

We also note that choosing the combinations of \( p_0, \gamma_1, \) and \( m_0 \) that resulted in the \( ANB(p_0) \) of the CUSUM being the largest (in 32 of the 40 cases) or second largest among the four methods (see Section 6.1) has the effect of diminishing the inequity in the \( SSANB \) comparisons due to the penalty incurred by the head-starts of the sets based methods.

We conclude this section with an argument as to why the \( SSANB \) is still the preferred metric for comparing methods that have head-start features. When practitioners design and implement the Sets, SHDA, or CUSCORE methods, they are unlikely to attempt to modify the procedure to remove the head-start. Subsequently, the threshold \( t \) will be chosen based on the usual \( ANB(p_0) \). Because convergence to the \( SSANB \) occurs rapidly (see Section 7.1.2), the \( SSANB \) is the most accurate measure of average run length performance unless it is probable that the shift will occur either prior to monitoring or during the very early stages of monitoring. In our view, comparing the unadjusted charts using the \( SSANB \) gives a more accurate representation of how the methods will perform in practice.
There are two other types of CUSUM charts that are applicable to the monitoring of a small incidence rate: the geometric CUSUM and the exponential CUSUM. The properties of the geometric CUSUM are discussed by Bourke [33, 34] and the exponential CUSUM was studied by Lucas [43] and by Vardeman and Ray [44]. Under mild conditions, the CUSUM chart based on the geometric distribution is equivalent to the Bernoulli CUSUM with a head-start [11, 34]. Just like the sets based methods, because the geometric CUSUM is based on geometric random variables, it implicitly assumes that a malformation was observed immediately prior to the onset of monitoring. This gives the geometric CUSUM a head start over the usual Bernoulli CUSUM. However, if we apply the same assumption (that a malformation occurred just prior to monitoring) to the Bernoulli CUSUM, the two charts can be constructed to be equivalent. Bourke [34] gives some empirical examples that demonstrate the equivalence.

Due to the close relationship between the Bernoulli and geometric CUSUM, Kenett and Pollak’s [14] comparison of the Sets method to a modified version of the geometric CUSUM is of particular interest. Kenett and Pollak modified the typical CUSUM scheme by not having the CUSUM statistic reset to zero after crossing the control limit. Instead, if the CUSUM statistic continues to increase after crossing the limit or if it descends but crosses the control limit again, it is considered to be additional evidence that the incidence rate has increased. Because the modified geometric CUSUM is not reset after an alarm, the two charts were designed to have equivalent false alarm rates (rather than equivalent values of $ANB(p_0)$). Kenett and Pollak compared the $ANB*(p_1)$ of the Sets method to the average number of births until the first crossing of the control limit of the modified geometric CUSUM and demonstrated that for $\gamma_1 = 6$ and $\gamma_1 = 7$, the $ANB(p_1)$ of the Sets method is 5 to 32% larger than the corresponding $ANB(p_1)$ of the modified geometric CUSUM. This suggests that the (unmodified) geometric CUSUM may be more efficient in detecting large increases in the incidence rate. However, the comparisons were made using Chen’s original design parameters [7] which were shown to be suboptimal for these types of comparisons [9, 17], and the scope of the study was limited only to 6 and 7-fold increases in the baseline incidence rate.

Bourke [33] notes that for very small values of $p_0$, the distribution of the geometric variables $X_i$ are approximately exponential with mean $(1 - p_0)/p_0$. Thus, the geometric CUSUM can be thought of as a discrete version of the exponential CUSUM. Hence, the Bernoulli CUSUM is similar to the geometric CUSUM and the geometric CUSUM is similar to the exponential CUSUM for small $p_0$. If we complete the syllogism, it can be argued that for small $p_0$, the Bernoulli CUSUM with a head-start should have performance
characteristics that are similar to the exponential CUSUM. This relationship between the Bernoulli and the exponential CUSUM is of interest because we can gain insight into how the Bernoulli CUSUM might compare to the Poisson CUSUM. Gan [45] compares the performance of exponential CUSUM and Poisson CUSUM charts. The key conclusion is that for detecting increases in the incidence rate the exponential CUSUM is more efficient and for detecting decreases in the incidence rate, the Poisson CUSUM is more efficient. This is because malformations are observed more frequently when the incidence rate increases, and each time a malformation is observed, the exponential CUSUM chart reacts immediately—whereas the Poisson CUSUM must wait until the monitoring interval is complete before it reacts to an increase in the incidence rate. When the rate decreases, the interarrival times lengthen and thus the exponential CUSUM takes longer to react because it has to wait for the occurrence of a malformation, whereas the Poisson CUSUM will update the chart as soon as the monitoring interval ends. Clearly, the difference between the Poisson and the exponential CUSUM charts depends on the length of the monitoring interval of the Poisson CUSUM. The disparity between the charts becomes more pronounced with larger Poisson monitoring intervals. Interestingly enough, if the size of the monitoring interval were so small (and the arrival of births so regular) that each interval contained only one birth, the Poisson process reduces to a sequence of Bernoulli trials.

7.5 Guidelines for practitioners

For the practitioner, one of the principle findings of our work is that the Bernoulli CUSUM is better than the Sets, SHDA, and CUSCORE methods at detecting increases in a small incidence rate. One may argue that the Sets and CUSCORE methods are preferable because they are relatively easy to design and implement. However, the superior performance of the CUSUM, coupled with the computational power available from modern computers certainly justifies the added complexity that may be required to design and implement a Bernoulli CUSUM monitoring scheme. Reynolds and Stoumbos [11] gave a number of tables for the design of CUSUM charts when the baseline incidence rate $p_0$ is 0.001 or larger. However, the incidence of a congenital malformation is often considerably smaller than 1 in 1,000. For this reason we include Tables 7.6 and 7.7 to aid in the design of CUSUM charts when $p_0$ is as small as 1 in 100,000.
Table 7.3: Number of births as a fraction of $ANB(p_0)$ at which $WANB(\tau_1 a)/SSANB(\gamma_1 a)$ falls and remains within the interval (0.999, 1.001). Eleven cases for the CUSUM were omitted because they were not computationally feasible.
Table 7.4: Thirteen instances when the $SSANB(\gamma)$ performance of the CUSCORE or the SHDA methods were better than the CUSUM in the smaller study with 23 cases.

<table>
<thead>
<tr>
<th>Method</th>
<th>$p_0$</th>
<th>$b_0$</th>
<th>$\gamma_1$</th>
<th>$\gamma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>CUSCORE</td>
<td>0.001</td>
<td>50,000</td>
<td>6</td>
<td>1.25, 1.5, 1.75</td>
</tr>
<tr>
<td>CUSCORE</td>
<td>0.0001</td>
<td>500,000</td>
<td>6</td>
<td>1.25, 1.5, 1.75</td>
</tr>
<tr>
<td>SHDA</td>
<td>0.01</td>
<td>25,000</td>
<td>6</td>
<td>1.25, 1.5, 1.75, 2.0</td>
</tr>
<tr>
<td>SHDA</td>
<td>0.005</td>
<td>50,000</td>
<td>6</td>
<td>1.25, 1.5, 1.75</td>
</tr>
</tbody>
</table>

Table 7.5: Values of $SSANB(3.5)$ for case 26 when the methods are adjusted to remove the effect of the head-starts of the sets based methods and when they are not adjusted.

<table>
<thead>
<tr>
<th></th>
<th>Sets</th>
<th>SHDA</th>
<th>CUSCORE</th>
<th>CUSUM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unadjusted</td>
<td>2654.57</td>
<td>2749.75</td>
<td>2620.12</td>
<td>2222.71</td>
</tr>
<tr>
<td>Adjusted</td>
<td>2654.57</td>
<td>2649.98</td>
<td>2620.12</td>
<td>2228.16</td>
</tr>
<tr>
<td>$p_0 \times 10^5$</td>
<td>$\gamma_1$</td>
<td>$\gamma_{1a}$</td>
<td>$m$</td>
<td>25</td>
</tr>
<tr>
<td>------------------</td>
<td>------------</td>
<td>---------------</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>100.0</td>
<td>1.5</td>
<td>1.49971</td>
<td>811</td>
<td>2931</td>
</tr>
<tr>
<td></td>
<td>(25002)</td>
<td>(50003)</td>
<td>(100037)</td>
<td>(250034)</td>
</tr>
<tr>
<td></td>
<td>2.0</td>
<td>2.00056</td>
<td>693</td>
<td>2171</td>
</tr>
<tr>
<td></td>
<td>(25012)</td>
<td>(50057)</td>
<td>(100014)</td>
<td>(250166)</td>
</tr>
<tr>
<td></td>
<td>3.0</td>
<td>3.00192</td>
<td>549</td>
<td>1465</td>
</tr>
<tr>
<td></td>
<td>(25047)</td>
<td>(50093)</td>
<td>(100055)</td>
<td>(250007)</td>
</tr>
<tr>
<td></td>
<td>5.0</td>
<td>5.00300</td>
<td>402</td>
<td>906</td>
</tr>
<tr>
<td></td>
<td>(25006)</td>
<td>(50015)</td>
<td>(100148)</td>
<td>(250166)</td>
</tr>
<tr>
<td>75.0</td>
<td>1.5</td>
<td>1.50059</td>
<td>1081</td>
<td>3907</td>
</tr>
<tr>
<td></td>
<td>(33350)</td>
<td>(66864)</td>
<td>(133384)</td>
<td>(333373)</td>
</tr>
<tr>
<td></td>
<td>2.0</td>
<td>2.00061</td>
<td>924</td>
<td>2895</td>
</tr>
<tr>
<td></td>
<td>(33340)</td>
<td>(66711)</td>
<td>(133341)</td>
<td>(335527)</td>
</tr>
<tr>
<td></td>
<td>3.0</td>
<td>3.00215</td>
<td>732</td>
<td>1953</td>
</tr>
<tr>
<td></td>
<td>(33341)</td>
<td>(66714)</td>
<td>(133503)</td>
<td>(333308)</td>
</tr>
<tr>
<td></td>
<td>5.0</td>
<td>5.00403</td>
<td>536</td>
<td>1209</td>
</tr>
<tr>
<td></td>
<td>(33384)</td>
<td>(66786)</td>
<td>(133640)</td>
<td>(333983)</td>
</tr>
<tr>
<td>50.0</td>
<td>1.5</td>
<td>1.49973</td>
<td>1622</td>
<td>5865</td>
</tr>
<tr>
<td></td>
<td>(50021)</td>
<td>(100008)</td>
<td>(200013)</td>
<td>(500138)</td>
</tr>
<tr>
<td></td>
<td>2.0</td>
<td>2.00066</td>
<td>1386</td>
<td>4344</td>
</tr>
<tr>
<td></td>
<td>(50026)</td>
<td>(100020)</td>
<td>(200107)</td>
<td>(500249)</td>
</tr>
<tr>
<td></td>
<td>3.0</td>
<td>2.99775</td>
<td>1099</td>
<td>2935</td>
</tr>
<tr>
<td></td>
<td>(50042)</td>
<td>(100104)</td>
<td>(200194)</td>
<td>(500423)</td>
</tr>
<tr>
<td></td>
<td>5.0</td>
<td>4.99518</td>
<td>805</td>
<td>1817</td>
</tr>
<tr>
<td></td>
<td>(50010)</td>
<td>(100118)</td>
<td>(200244)</td>
<td>(500858)</td>
</tr>
<tr>
<td>25.0</td>
<td>1.5</td>
<td>1.49975</td>
<td>3244</td>
<td>11731</td>
</tr>
<tr>
<td></td>
<td>(10018)</td>
<td>(20017)</td>
<td>(400030)</td>
<td>(100060)</td>
</tr>
<tr>
<td></td>
<td>2.0</td>
<td>1.99942</td>
<td>2773</td>
<td>8695</td>
</tr>
<tr>
<td></td>
<td>(100027)</td>
<td>(200007)</td>
<td>(40009)</td>
<td>(100018)</td>
</tr>
<tr>
<td></td>
<td>3.0</td>
<td>3.00029</td>
<td>2197</td>
<td>5867</td>
</tr>
<tr>
<td></td>
<td>(100053)</td>
<td>(200044)</td>
<td>(40004)</td>
<td>(100150)</td>
</tr>
<tr>
<td></td>
<td>5.0</td>
<td>5.00114</td>
<td>1609</td>
<td>3632</td>
</tr>
<tr>
<td></td>
<td>(100021)</td>
<td>(200051)</td>
<td>(40007)</td>
<td>(100026)</td>
</tr>
<tr>
<td>10.0</td>
<td>1.5</td>
<td>1.50010</td>
<td>8109</td>
<td>29322</td>
</tr>
<tr>
<td></td>
<td>(25013)</td>
<td>(50012)</td>
<td>(100017)</td>
<td>(250031)</td>
</tr>
<tr>
<td></td>
<td>2.0</td>
<td>2.00022</td>
<td>6931</td>
<td>21730</td>
</tr>
<tr>
<td></td>
<td>(25012)</td>
<td>(50023)</td>
<td>(100013)</td>
<td>(250023)</td>
</tr>
<tr>
<td></td>
<td>3.0</td>
<td>2.99997</td>
<td>5493</td>
<td>14670</td>
</tr>
<tr>
<td></td>
<td>(250003)</td>
<td>(500053)</td>
<td>(100022)</td>
<td>(250035)</td>
</tr>
<tr>
<td></td>
<td>5.0</td>
<td>5.00077</td>
<td>4023</td>
<td>9084</td>
</tr>
<tr>
<td></td>
<td>(250050)</td>
<td>(500124)</td>
<td>(100030)</td>
<td>(250093)</td>
</tr>
</tbody>
</table>

Table 7.6: Values of $w$ for the Bernoulli CUSUM with the corresponding actual $ANB_b(p_0)$ underneath in parentheses. Note that $h = w/m$. 
\[
\begin{array}{|c|c|c|c|c|c|c|c|}
\hline
p_0 \times 10^5 & \gamma_1 & \gamma_{1a} & m & 25 & 50 & 100 & m_0 \\
\hline
1.5 & 1.500 & 10812 & 39096 & 51074 & 64820 & 85206 & 101859 & 111945 \\
& & & (333335) & (666671) & (1333357) & (3333399) & (6666726) & (10000106) \\
2.0 & 2.000 & 9242 & 28978 & 36382 & 44420 & 55701 & 64552 & 69810 \\
& & & (333354) & (666687) & (1333336) & (3333466) & (6666785) & (10000068) \\
7.5 & 3.000 & 7324 & 19561 & 23678 & 27964 & 33876 & 38419 & 41090 \\
& & & (333365) & (666731) & (1333459) & (3333849) & (6667226) & (10000455) \\
5.0 & 4.999 & 5365 & 12116 & 14400 & 16566 & 19586 & 21895 & 23238 \\
& & & (333399) & (666721) & (1333348) & (3333899) & (666894) & (10002051) \\
5.0 & 5.000 & 8047 & 18172 & 21599 & 24848 & 29377 & 32841 & 34854 \\
& & & (500013) & (10000105) & (20000190) & (5000451) & (10001291) & (15000347) \\
5.0 & 5.000 & 8047 & 18172 & 21599 & 24848 & 29377 & 32841 & 34854 \\
& & & (500013) & (10000105) & (20000190) & (5000451) & (10001291) & (15000347) \\
2.5 & 1.500 & 32437 & 117300 & 153239 & 194483 & 255653 & 305622 & 335887 \\
& & & (2500019) & (5000017) & (10000006) & (25000060) & (5000232) & (75000211) \\
5.0 & 5.000 & 16094 & 54931 & 54319 & 49698 & 58756 & 65684 & 69711 \\
& & & (2500018) & (5000102) & (10000105) & (25000018) & (5000501) & (7500267) \\
5.0 & 5.000 & 16094 & 54931 & 54319 & 49698 & 58756 & 65684 & 69711 \\
& & & (2500017) & (5000126) & (10000017) & (25000353) & (50000012) & (75000559) \\
\hline
\end{array}
\]

Table 7.7: Values of \( w \) for the Bernoulli CUSUM with the corresponding actual \( ANB_b(p_0) \) underneath in parentheses. Note that \( h = w/m \).
Tables 7.6 and 7.7 contain values of $m$ (the reciprocal of the reference value $\delta$), $w$ (the number of states), and the associated values of $ANB_0(p_0)$ (underneath in parentheses) for various combinations of the incidence rate, $p_0$, the shift size, $\gamma_1$, and the target in-control average number of malformations until signal, $m_0$, for $p_0 \leq 0.001$. The control limit $h$ is given simply by $w/m$. To implement the CUSUM, a reliable estimate of $p_0$ should be obtained. The desired value of $m_0$ and a shift size $\gamma_1$ for which the chart will be optimal must also be specified. Then the values of $p_0, \gamma_1,$ and $m_0$ from Table 7.6 or 7.7 that most closely resemble the desired values can be identified, along with the corresponding design parameters $m$ and $w$. The CUSUM statistic $B_j$ is given by

$$B_0 = 0$$
$$B_j = \max \left( 0, B_{j-1} + V_j - \frac{1}{m} \right) \quad j = 1, 2, \ldots ,$$

where $V_j = 1$ indicates the presence of a malformation and 0 indicates its absence. An alarm is signaled when $B_j \geq h$. Because the values of $m$ are so large, this approach is easily affected by rounding error. We can avoid this by equivalently defining the CUSUM statistic so that it only takes on integer values as follows:

$$B_0 = 0$$
$$B_j = \max \left( 0, B_{j-1} + V_j - m - 1 \right) \quad j = 1, 2, \ldots ,$$

and we signal an alarm when $B_j \geq w$.

Since it may be tedious to evaluate the CUSUM statistic after each birth, we can instead update the CUSUM statistic after each incident is observed, provided we count the number of births since the last incident. This is possible because there is no loss of information in only observing the number of births between incidents and because the CUSUM can only signal after the observance of a malformation. As before, let $X_i$ equal the number of (normal) births observed between (but not including) incidents $i$ and $i + 1$. We can then equivalently define the CUSUM statistic as

$$B_0 = 0$$
$$B_i = \max \left( 0, B_{i-1} - X_i + m - 1 \right) \quad i = 1, 2, \ldots \,$$

and signal an alarm when $B_i \geq w$.

Depending upon the context in which the CUSUM is used, when an alarm is signaled, it may be useful to not reset the chart but to continue monitoring to see if the CUSUM statistic remains above the control
limit or if additional upcrossings of the control limit occur. Such behavior would provide additional evidence that the alarm was real and that an increase in the baseline incidence rate has indeed occurred. However, this approach does make it increasingly difficult to calculate the false alarm rate that is needed to determine the control limit [14]. Since it is desirable to have some type of prespecified decision rule that can be used to sound an alarm, not resetting the chart when the CUSUM statistic crosses the control limit makes the formulation of a decision rule more complicated. Regardless of whether the chart is reset, any alarm should provide the impetus for additional epidemiological investigation.

In some monitoring situations, it may not be possible to directly observe a sequence of Bernoulli trials. Depending on how the data are observed, it may be determined that the geometric, exponential, or Poisson CUSUM would be best suited to the surveillance application. Practitioners can design the appropriate chart by referring to Bourke [33] for the geometric CUSUM, Gan [45] and Vardeman and Ray [44] for the exponential CUSUM, and Ewan and Kemp [15] or Lucas [43] for the Poisson CUSUM. If a head-start (also known as a fast initial response) is desired, any type of CUSUM chart can be modified to have a head-start by initializing the CUSUM statistic at a value above zero. Bourke [33] and Lucas [43] gave tables of design parameters for designing geometric and Poisson CUSUM charts with head-start features, respectively. Even though the sets based methods have head-starts, we do not recommend their use as they are inherently less efficient than CUSUM charts because of their “direct loss of information owing to the dichotomization of data” [4].
Chapter 8

Conclusions

The surveillance of birth defects continues to be of interest and concern to epidemiologists. Since its introduction in 1978, the Sets method has received considerable attention. Even recently, Grigg and Farewell [27, 46] adapted the Sets method to monitor binary surgical outcomes where the baseline probability of failure differs for each patient, and Carpenter [47] incorporated financial considerations with the Sets method to make decisions regarding epidemics in cattle populations.

8.1 In search of the most equitable comparison

Various comparisons involving the Poisson CUSUM and the Sets, CUSCORE, and SHDA methods have been published, and yet in most cases, key issues regarding the assumptions about when the shift occurs, the criteria for choosing the design parameters, the metric used to compare the methods, and the head-start features of the methods have been overlooked. In particular, making the assumption that the shift in the incidence rate takes place prior to monitoring and then comparing methods with the out-of-control initial state $ARL$ confers an unfair advantage on methods that have head-starts. In addition, if the time of the shift does not necessarily occur before the onset of monitoring, the fact that the shift is not likely to occur right after a malformation must be accounted for. Bourke [34] aptly discusses a number of these same issues with respect to analyzing the performance of the geometric CUSUM. The concerns we have raised with the previous comparisons involving the sets based methods are substantiated by this observation from Sonesson and Bock [4]:
“In many of the papers [which deal] with the inferential aspects [of the methods] correctly, the lack of a proper statistical evaluation of the methods suggested is evident. Usually, the only measures that are considered are $ARL^0$ and $ARL^1$. However, in public health surveillance the event to be detected is not likely to occur at the same time as the surveillance starts. This means that $ARL^1$ is not a suitable measure of evaluation. Instead, other types of measure should be used, taking into account also possible later changes, since the performance of a surveillance method depends on the time of the change.”

In the comparisons presented in Part I, we attempted to address the four concerns mentioned in Chapter 2 as far as it was computationally feasible. First, our choice of the SSANB as the principal metric of comparison only assumes that the chart has not yet signaled and that the shift takes place after the chart has reached the steady-state distribution. The simplifying assumption that a false alarm has not occurred prior to convergence to the steady-state distribution does not compromise our ability to equitably compare the various methods. And because the convergence to the steady-state distribution occurs rapidly, assuming that the shift takes place after convergence is not unreasonable. Alternatively, using a metric that averages the chart performance over a specified distribution of shift times (see equation (7.1)) is also possible, albeit computationally intensive. Second, the advantage of the head-starts of the sets based methods was essentially nullified by evaluating the methods using the SSANB. We did this so as not to disadvantage the Bernoulli CUSUM. As discussed in Section 7.3.2, the extent to which comparing the charts using the SSANB was a disadvantage to the sets based methods appears to be minimal. Third, we chose the design parameters to be optimal according to the same criteria (minimizing the worst case $ARL$). In Section 7.3.1 we also considered the comparisons of the methods when the design parameters were chosen to minimize the initial state $ARL$, and found that the conclusions remain virtually unchanged. And fourth, we examine the performance of the methods across a range of possible shift sizes.

In addition, we only considered combinations of the baseline incidence rate, the target in-control $ARL$, and the shift size that resulted the closest possible matches of the in-control $ARL$’s of the four methods. Wherever possible, the cases were selected so that the CUSUM had the largest $ANB(p_0)$ among the four methods. The vast majority of the $ARL$ calculations were performed exactly, and simulation error was accounted for when simulation was necessary.
8.2 Summary of results

In Chapter 6 we conclusively demonstrate that the Bernoulli CUSUM chart outperforms the Sets, CUSCORE, and SHDA methods under a wide variety of circumstances. We summarize the key results of the comparisons below:

- The CUSUM always has the lowest worst case $ARL$ at the shift for which the chart was designed (Moustakides’ result).

- The steady-state average number of births from shift until signal, evaluated at the shift for which the chart was designed ($SSANB(\gamma_1)$) is lowest for the CUSUM among the forty diverse cases that were considered.

- The $SSANB$ performance of the CUSUM across a range of hypothetical shifts is almost uniformly better than the other three methods, with the exception of only 13 out of 1160 instances in which the CUSCORE was slightly better than the CUSUM. This occurred when the shift for which the chart was designed was relatively large ($\gamma_1 \geq 4$) and the actual shift was small ($\gamma \leq 1.75$).

- In all forty cases that were considered, the CUSUM has the best performance in terms of area underneath the $SSANB$ curve ($\int_{1.25}^{7.75} \log(SSANB(\gamma))d\gamma$).

- In addition, examination of the $SSANB$ profiles suggests that the shift size for which the charts are optimized seems to most influence the difference in performance between the four methods. The most noteworthy difference is that if the charts are designed for a small shift and a large one actually occurs, the CUSUM and the CUSCORE substantially outperform the Sets and the SHDA Methods.

- Multiple regression models demonstrate that the improvement in performance of the CUSUM over the CUSCORE does not depend on $p_0$, but is affected by $\gamma_1$ and $m_0$. In particular, maximal improvement of the CUSUM over the CUSCORE occurs when the target $ANM(p_0)$ is about 640 incidents.

- If we only consider the improvement in $SSANB$ performance at the shift for which the charts were designed, the improvement of the CUSUM over the CUSORE is highest when the charts are designed for small shift sizes, and this improvement decreases linearly as the shift size increases.

- With regard to the improvement in terms of area under the $SSANB$ curve, the CUSUM shows maximal improvement over the CUSCORE when the charts are designed to detect a 3.7 fold increase in the baseline incidence rate.
8.3 Final remarks

We also demonstrate that, in many respects, the four methods can be ranked in terms of the efficiency with which they accumulate information and react to a shift in the incidence rate. In particular, the Bernoulli CUSUM is best, followed by the CUSCORE, then the Sets method, and last of all, the SHDA method. The SHDA method is worst because of the possibility that at least two flags must be raised in order to signal—thereby incurring extra delay in detecting the shift. Although the delay in the Sets method is less pronounced, requiring that \( n_s \) sets in a row be less than the threshold forces the counter \( S_i \) to reset whenever a set exceeds the threshold \( t_s \), thereby discarding all previous information accumulated in the chart. Among the sets based methods, the CUSCORE is the most efficient because a set that exceeds the threshold does not necessarily reset the chart, but rather moves it down one step, allowing the chart to accumulate information over a longer period of time. But ultimately, the dichotomization of sets into those that exceed the threshold and those that do not reduces the amount of information contained in the chart, and this is why the CUSUM is the most efficient of the four methods we consider.

The Bernoulli CUSUM is one of the best statistical tools available for monitoring a small incidence rate (proportion). It can readily be adapted to monitor for decreases in the incidence rate or to simultaneously monitor for increases and decreases in the incidence rate [11]. Although the context of our discussion has been the surveillance of congenital malformations, the Bernoulli CUSUM is certainly applicable to a wide range of applications in monitoring and surveillance.
Part II

Risk-adjusted monitoring of clinical outcomes
Chapter 9

Introduction

In Part I we focused on monitoring the incidence rate of a rare health event where the probability of the event was assumed to be constant across all individuals who were monitored. In Part II, we consider the situation where the probability of the outcome of interest varies for each subject (or object) that is plotted in the control chart. We develop methods for monitoring the survival time outcomes of subjects who each have a potentially different survival distribution. The motivating example for this work is the monitoring of surgical outcomes, where each patient has a different risk of dying due to their unique health history. These methods are especially pertinent in light of the increasing awareness of the need to implement statistical process control to improve the quality of health care [48, 49, 50, 51]. There are a number of issues that must be carefully considered in order to accurately monitor surgical outcomes [52, 53, 54, 55]. Foremost among them is the need to account for the fact that each patient has a different risk of failure due to his/her age, gender, medical histories, etc. This heterogeneity among patients is often referred to as the patient “mix.” In addition, it may also be of interest to account for risk factors that are associated with the specific type of surgical procedure that will be performed. Incorporating the risk information that is known prior to the procedure into a control chart is known as “risk-adjustment.” If a control chart does not adjust for the patient mix, the outcomes plotted on the chart become confounded with the preexisting risk factors. This makes it nearly impossible to clearly identify special causes related to the surgical procedure and, consequently, the control chart would have little value as a tool for quality improvement.

Steiner et al. [56] gave a compelling example of the consequences of not risk-adjusting for the patient mix. Experienced surgeons typically operate on patients that are in worse condition than patients that are referred to training surgeons. Steiner et al. demonstrated that not accounting for the patient mix assigned to
a given surgeon would lead to the erroneous conclusion that the quality of surgeries performed by the trainees is better than that of the experienced surgeons. The authors aptly note that the erroneous conclusions would have resulted in wasted time and effort in searching for a special cause to explain the “good” performance of the training surgeons and the “poor” performance of the experienced surgeons. Clearly risk-adjustment is essential when monitoring a health care procedure performed on a wide variety of patients.

A number of risk-adjusted (RA) control chart methods have been proposed [46, 51]. Among these are the variable life-adjusted display (VLAD) [57] and the equivalent cumulative RA mortality (CRAM) chart [58], the RA Shewhart $p$-chart [59, 60], the RA Sets method [27], the RA resetting sequential probability ratio test (RSPRT) [61, 62], and the RA Bernoulli CUSUM chart [56, 63]. While the VLAD and the CRAM charts provide intuitive graphical displays, they lack meaningful control limits [51] used to formally indicate the improvements or degradation in the quality of a surgical procedure. It is possible to use both a VLAD chart and a RA Bernoulli CUSUM chart together to monitor a process in order to combine the ease of interpretation afforded by the VLAD chart with the formal statistical inference provided by the RA Bernoulli CUSUM chart [64]. Grigg and Farewell [46] concluded that the RA Bernoulli CUSUM chart is more efficient than the RA Shewhart $p$-chart, even when the increase in the odds of failure is large. They also suggested that the RA Sets method is better than the RA Bernoulli CUSUM chart at detecting large increases in the odds of death, especially when the in-control $ARL$ is small. However, since these two methods were compared using the initial state $ARL$, there is an implied assumption that the shift in the odds of death has occurred prior to the onset of monitoring, thereby favoring the RA Sets method because it has a head-start over the RA Bernoulli CUSUM chart. Furthermore, the Sets method is inherently inefficient because it effectively reduces the amount of information that is included in the chart [4]. In Part I we demonstrated that when no risk-adjustment is used, the Bernoulli CUSUM chart uniformly outperforms the Sets method for each of the cases that were considered. This suggests that the RA Bernoulli CUSUM chart should outperform the RA Sets method—although further study may be warranted. One of the weaknesses of the RA RSPRT is that the chart can accumulate too much “credit” [46] or “inertia” [65], thereby reducing its responsiveness in detecting a decrease in quality that occurs after a period of higher quality. Interestingly enough, Spiegelhalter, one of the original proponents of the RA RSPRT, now prefers the RA Bernoulli CUSUM chart over the RA RSPRT [52], presumably because the RA RSPRT is apt to accumulate too much credit.

As a tool for monitoring binary surgical outcomes, we prefer the RA Bernoulli CUSUM chart because 1) it does not build up credit and thus retains sensitivity to shifts in the mortality rate, and 2) because it is based upon Page’s [10] original CUSUM scheme which has been shown to have optimality properties [38].
The RA Bernoulli CUSUM chart has been recently applied in a number of contexts [66, 67, 68, 69].

A common measure of clinical performance is the 30-day mortality rate, which is the percentage of patients who die within 30 days following the procedure. Perhaps as a consequence, the risk-adjusted charts that are discussed in the literature are almost exclusively designed to monitor binary outcomes [27, 46, 54, 56, 57, 58, 59, 60, 62, 70]. However, it would be more informative to monitor the patient survival time following the procedure. This of course would require the censoring of observations when 1) the patient has not yet died at the time when the control chart is created, 2) the patient dies due to a cause unrelated to the surgery, or 3) it is only known that a patient was still alive at a certain time because information about the patient is no longer available. To this end, we propose a risk-adjusted survival time CUSUM chart (RAST CUSUM) that is designed to monitor a continuous, right-censored time-to-event variable. The statistical performance of the new RAST CUSUM chart will be compared to the RA Bernoulli CUSUM chart.

The prospective monitoring of a binary variable is straightforward. For example, to monitor the 30-day mortality rate for a surgical procedure, the binary outcome for each patient is plotted 30 days after surgery in the order which the patients received the surgery. However, the implementation of a prospective monitoring scheme for a right-censored time-to-event variable is not as obvious, since the order in which the death times or censored times are observed may not coincide with the order in which the operations were performed—especially if survival times longer than 30 days are considered. We address this and other issues regarding the use of the RAST CUSUM chart for prospective monitoring in greater detail in Section 13.2.
Chapter 10

Description of the risk-adjusted CUSUM charts

10.1 Risk-adjustment

One of the primary challenges in implementing a risk-adjusted monitoring system is the development of models which can accurately predict the risk of death following the operation for each patient, based upon risk variables which are measured prior to the operation. Steiner et al. [56] refer to this as the “pre-operative risk of surgical failure.” Thus, the risk-adjustment models reflect the historical risk of surgical failure (i.e. preoperative risk) by accounting for the patient mix. A common approach is to use an index that is a weighted composite of the factors that are deemed most likely to influence preoperative risk associated with a surgical procedure. For example, the Parsonnet score [71] and the EuroSCORE [72] are commonly used estimate the risk for cardiac patients. The APACHE III score [73] has been used for RA monitoring of outcomes from intensive care units [60]. Alemi et al. [59] discussed six other indexes designed to assess preoperative risk. Table 10.1 shows the risk factors and the corresponding weights that compose the Parsonnet score. The fact that it incorporates twenty-four different factors speaks to the complexity of accurately measuring the preoperative risk for any given patient. While these indexes do have limitations [74], such as the availability of valid databases and the lack of consensus among clinicians on how to assess risk [75, 76], the overarching need to include risk-adjustment as part of the monitoring procedures still remains.

When monitoring a binary outcome, the preoperative risk for each patient can be estimated using a variety of models. Specifically, let \( p_i(\beta) \) represent the failure rate for patient \( i \). Assume that

\[
p_i(\beta) = g(\beta, u_i) \tag{10.1}
\]
<table>
<thead>
<tr>
<th>Parsonnet factor</th>
<th>Weight(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Patient factors</strong></td>
<td></td>
</tr>
<tr>
<td>Female sex</td>
<td>1</td>
</tr>
<tr>
<td>Left ventricular ejection fraction 30-49%</td>
<td>2</td>
</tr>
<tr>
<td>Preoperative balloon pump</td>
<td>2</td>
</tr>
<tr>
<td>Weight $\geq$ 1.5 times ideal</td>
<td>3</td>
</tr>
<tr>
<td>Diabetes</td>
<td>3</td>
</tr>
<tr>
<td>Hypertension (systolic pressure $\geq$ 140 mm Hg)</td>
<td>3</td>
</tr>
<tr>
<td>Left ventricular ejection fraction $&lt; 30%$</td>
<td>4</td>
</tr>
<tr>
<td>Age 70-74</td>
<td>7</td>
</tr>
<tr>
<td>Dependent on dialysis</td>
<td>10</td>
</tr>
<tr>
<td>Emergency due to failure in cardiac catheter laboratory</td>
<td>10</td>
</tr>
<tr>
<td>Age 75-79</td>
<td>12</td>
</tr>
<tr>
<td>Age $\geq 80$</td>
<td>20</td>
</tr>
<tr>
<td>Catastrophic state</td>
<td>30</td>
</tr>
<tr>
<td><strong>Procedural factors</strong></td>
<td></td>
</tr>
<tr>
<td>CABG at time of valve surgery</td>
<td>2</td>
</tr>
<tr>
<td>Valve replacement:</td>
<td></td>
</tr>
<tr>
<td>Tricuspid valve</td>
<td>3</td>
</tr>
<tr>
<td>Aortic valve (gradient $\leq$ 120 mm Hg)</td>
<td>5</td>
</tr>
<tr>
<td>Mitral valve (PASP $&lt; 60$ mm HG)</td>
<td>5</td>
</tr>
<tr>
<td>Congenital heart disease</td>
<td>5</td>
</tr>
<tr>
<td>Left ventricular aneurysm</td>
<td>5</td>
</tr>
<tr>
<td>First reoperation</td>
<td>5</td>
</tr>
<tr>
<td>Valve replacement:</td>
<td></td>
</tr>
<tr>
<td>Aortic valve (gradient $&gt; 120$ mm Hg)</td>
<td>7</td>
</tr>
<tr>
<td>Mitral valve (PASP $\geq 60$ mm HG)</td>
<td>8</td>
</tr>
<tr>
<td>Second or subsequent reoperation</td>
<td>10</td>
</tr>
<tr>
<td>Heart transplantation</td>
<td>15</td>
</tr>
</tbody>
</table>

Table 10.1: Parsonnet risk factors. To calculate the Parsonnet score, add all weights that apply to the patient and the type of operation—for example, a female patient having tricuspid valve replacement has an estimated risk of $1+3=4\%$. CABG = coronary artery bypass grafting and PASP = pulmonary artery systolic pressure. The content of this table was given by Poloniecki et al. [58].

where $\mathbf{u}_i$ is a vector of covariates that reflect the risk factors for patient $i$ and $\mathbf{\beta}$ is a corresponding vector of parameters. For example, $\mathbf{\beta}$ could contain the parameters used in a logistic regression model, and $g(\cdot)$ would correspond to the inverse logit function. The model can include several covariates or simply a single descriptive index, such as the Parsonnet score.

In what follows we make the simplifying assumption that the risk-adjustment model built from historical data is, in fact, an accurate model and that the regression parameter vector $\mathbf{\beta}$ is known. Steiner et al. [56] investigated the impact of estimating the parameter vector $\mathbf{\beta}$ by fitting the logistic regression model to bootstrap samples of the historical data set and then plotting the resulting RA Bernoulli CUSUM
paths of the “prospective” data set for each bootstrap sample. The results suggest that while estimating the regression parameters does impact the path of the RA Bernoulli CUSUM, the impact is not so severe as to alter the general conclusions. Jensen et al. [29] gave a review of non risk-adjusted control chart literature that discusses the impact of assuming certain parameters (e.g. $\beta$) are known, when, in fact, they are estimated from historical data.

10.2 The risk-adjusted Bernoulli CUSUM chart

In this section we base our description of the RA Bernoulli CUSUM chart on that given by Steiner et al. [56, 63]. We have made changes to the notation and inserted additional commentary to clarify the exposition. The basic form of the CUSUM chart is given by

\begin{equation}
Z_0^+ = 0 \\
Z_i^+ = \max(0, Z_{i-1}^+ + W_i) \quad i = 1, 2, \ldots
\end{equation}

where $Z_i^+$ is the CUSUM statistic, $W_i$ is the CUSUM score, and $i$ indexes the patients. An alarm is signaled if $Z_i^+ > h^+$. The superscript “+” is used to denote that the CUSUM statistic is bounded below by zero and can move upward toward the control limit, $h^+$. The CUSUM score, $W_i$, is the logarithm of the likelihood ratio corresponding to the random variable being monitored. The likelihood ratio depends on the in-control and the nominal out-of-control values of the parameter interest. The difference between the in-control and out-of-control values of this parameter should correspond to a meaningful (and interpretable) change in the quality of the process that we wish to be able to detect quickly. Moustakides [38] proved that CUSUM scores based on the likelihood ratio are optimal in the sense that, among the class of all monitoring schemes with the same in-control $ARL$, the CUSUM chart will have the smallest out-of-control $ARL$ when the shift in quality takes place at the worst possible time and when the chart is in a state that is furthest from signal. Depending on the chosen value of the out-of-control parameter, the upper CUSUM chart can be constructed to detect either increases or decreases in quality. For example, Steiner et al. [56] uses the upper CUSUM chart to detect increases in the mortality rate (a decrease in quality). A lower CUSUM chart that is bounded above by zero which can move downward toward a control limit is defined by

\begin{equation}
Z_0^- = 0 \\
Z_i^- = \min(0, Z_{i-1}^- - W_i) \quad i = 1, 2, \ldots
\end{equation}
with an alarm signaled if \( Z_i^- < h^- \). Like the upper CUSUM chart, the lower CUSUM chart may be used to detect either increases or decreases in process quality. The lower CUSUM chart is most likely to be used as part of a two-sided control scheme where the lower CUSUM chart is used in combination with an upper CUSUM chart in order to simultaneously monitor for both increases and decreases in quality. For instance, Steiner et al. \cite{56} used the upper CUSUM chart to monitor for increases in the mortality rate and the lower CUSUM chart to simultaneously monitor for decreases in the mortality rate. For a two-sided control scheme, the values of \( W_i \) in equations (10.2) and (10.3) charts would differ because the nominal out-of-control parameter value used to construct the likelihood ratio would depend on whether the chart was designed to detect an increase or decrease in quality. In this work, we focus primarily on the use of the upper CUSUM chart, since the implementation of the lower CUSUM chart is a straightforward extension of the upper CUSUM chart.

To motivate the construction of the RA Bernoulli CUSUM chart, let us first consider the regular (unadjusted) Bernoulli CUSUM chart that would be used to monitor a process where the failure rate is constant across all patients. Let \( p_0 \) represent the in-control failure rate and let \( p_1 \) represent the nominal out-of-control value of the failure rate. We begin by defining the Bernoulli random variables

\[
    Y_i = \begin{cases} 
        1 & \text{if patient } i \text{ dies} \\
        0 & \text{if patient } i \text{ survives} 
    \end{cases} \quad i = 1, 2, \ldots . 
\]

(10.4)

Letting \( f(y|p) \) denote the mass function of \( Y \), the scores \( W_{i}^{B} \) are given by

\[
    W_{i}^{B} = \log \left\{ \frac{f(y_i|p_1)}{f(y_i|p_0)} \right\} = \log \left\{ \frac{p_1^{y_i} (1-p_1)^{1-y_i}}{p_0^{y_i} (1-p_0)^{1-y_i}} \right\},
\]

(10.5)

where the superscript \( B \) indicates the score was derived using the Bernoulli distribution. The Bernoulli CUSUM chart is optimally designed to detect a change in the failure rate from \( p_0 \) to \( p_1 = \gamma_1 p_0 \). However, when risk-adjustment is required, the underlying value of \( p_i \) differs for each patient and is estimated using the model shown in equation (10.1). Because the predicted probability of failure for a high risk patient could be relatively large, a shift from \( p_i \) to \( \gamma_1 p_i \) may not make sense. For example, suppose \( \gamma_1 = 2 \) and \( p_i = 0.6 \). Obviously, it is nonsensical to consider a shift in the probability of failure from 0.6 to 1.2. However, we can circumvent this problem if the RA Bernoulli CUSUM chart is designed to detect a shift in the odds of patient \( i \) dying within a certain period of time following the operation (typically 30 days). The odds of
death is given by:

\[ O_i = \frac{P(\text{death})}{P(\text{survival})} = \frac{p_i}{1 - p_i} \quad . \tag{10.6} \]

Thus, in our previous numerical example, a two-fold increase in the odds of failure would correspond to a shift in \(p_i\) from 0.60 to 0.75. In more general terms, the inverse of the odds function, \(O/(1 + O)\), is a monotonic, increasing function that maps any shift in the odds of death to a corresponding shift in the probability of failure that lies within the unit interval. Conveniently, the odds of failure is a metric that is commonly used and understood by medical practitioners. The chart is designed to optimally detect a step-shift in the odds of death from \(O_{i0}\) to \(O_{i1} = \xi_1 O_{i0}\). We define the surgical process to be in control if the observed odds of death following the operation coincides with the odds of death predicted by the risk-adjustment model in equation (10.1). To detect increases in the death rate relative to the predicted mortality rate, we pick \(\xi_1 > 1\), and to detect decreases we choose \(0 < \xi_1 < 1\). A common choice for \(\xi_1\) is 2, corresponding to a doubling in the odds of failure.

To create the scores \((W_i^R)\) based on 30-day mortality, we define

\[ p_{i0} = \text{probability of death of patient } i \text{ within 30 days when the process is in control, and} \]
\[ p_{i1} = \text{probability of death of patient } i \text{ within 30 days when the process is out of control.} \]

We assume that when the process is in control,

\[ p_{i0} = p_i \quad . \tag{10.7} \]

Therefore, the nominal out-of-control odds of failure are given by

\[ O_{i1} = \frac{p_{i1}}{1 - p_{i1}} = \frac{\xi_1 p_{i0}}{1 - p_{i0}} \quad . \tag{10.8} \]

Then, solving equation (10.8) for \(p_{i1}\) in terms of \(\xi_1\) and \(p_{i0}\) gives

\[ p_{i1} = \frac{\xi_1 p_{i0}}{1 - p_{i0} + \xi_1 p_{i0}} \quad . \tag{10.9} \]

Substituting equations (10.7) and (10.9) into equation (10.5) with some algebra gives the following RA
Bernoulli CUSUM score:

\[ W_i^B(y_i) = \log \left\{ \frac{f(y_i|p_{i1})}{f(y_i|p_{i0})} \right\} = \log \left\{ \frac{\xi_i}{1 - p_i + \xi_i p_i} \right\}. \]  \hspace{1cm} (10.10)

We distinguish between the nominal shift in the odds of failure, \( \xi_1 \) and any arbitrary shift that might occur, \( \xi \). Therefore, when the process is in control, \( P(Y_i = 1) = p_i \), and when the odds of failure shifts from \( O_i \) to \( \xi O_i \), we have \( P(Y_i = 1) = \xi p_i/(1 - p_i + \xi p_i) \).

The control limits \( h_B^+ \) and/or \( h_B^- \) are determined by the desired in-control ARL. The ARL depends on the distribution of the covariates and the Bernoulli distribution of the patient outcomes conditional on the covariate. The ARL may be simulated or calculated approximately using the Markov chain approach. We discuss this in more detail in Chapter 11.

### 10.3 The risk-adjusted survival time CUSUM chart

The strategy for developing the RAST CUSUM chart is as follows. To account for the heterogeneity among patients, a survival time distribution will be predicted for each patient using an accelerated failure time (AFT) regression model. The predicted density and survival functions from the AFT model will then be used to construct a likelihood ratio that will be used for the scores in the RAST CUSUM chart. The method will account for right-censored observations, i.e., those patients that survive at least until an observed censoring time.

#### 10.3.1 Constructing the risk-adjusted likelihood ratio

We begin by considering the likelihood function for a right-censored, continuous time-to-event variable. Let \( X_i \) represent the survival time for patient \( i \) with survival function \( P(X_i > x_i) = S(x_i, \theta_i) \) and density \( f(x_i, \theta_i) \) where \( \theta_i \) is a vector of parameters that characterizes the survival model for patient \( i \). The time of death for some patients will not be observed. Rather, it will only be known that a patient survived at least until a random time \( C_i \). Alternatively, censoring times may be fixed in advance, so that each patient has a potentially different (but fixed) censoring time \( c_i \). There may even be a common censoring time \( c' \) for all patients—as would be the case if patients were only observed, say, for 30 days after the procedure (and if no patients were censored prior to 30 days). Regardless of whether the censoring mechanism is fixed or
random, the data are observed in pairs \((T_i, \delta_i)\) where

\[
T_i = \min(X_i, C_i) \quad \text{and} \quad \delta_i = \begin{cases} 
1 & \text{if } X_i \leq C_i \\
0 & \text{if } X_i > C_i 
\end{cases}
\]

(10.11)

For each of the three types of censoring mentioned above, the likelihood function for a single observation \((t_i, \delta_i)\) is given by

\[
L(\theta_i | t_i) = [f(t_i, \theta_i)]^{\delta_i} [S(t_i, \theta_i)]^{1-\delta_i}.
\]

(10.12)

Discussions of the derivation of the likelihood in equation (10.12) were given by Kalbfleisch and Prentice [77] and by Lawless [78]. We can construct the CUSUM scores to detect a change from \(\theta_i = \theta_i^0\) to \(\theta_i = \theta_i^1\). We discuss the nature of this shift in more detail in Section 10.3.3. The CUSUM score for patient \(i\) is then given by

\[
W_i = \log \left[ \frac{L(\theta_i^1 | t_i)}{L(\theta_i^0 | t_i)} \right]
\]

(10.13)

where \(L(\cdot)\) is given by equation (10.12). Then \(W_i\) can be used to sequentially calculate the CUSUM statistic as described by equations (10.2) and (10.3).

The control limits for the RAST CUSUM chart are chosen to achieve a desired in-control ARL. The ARL of the RAST CUSUM chart depends on the distribution of the covariate, the distribution of the survival and censor times conditional on the value of the covariate, and, if the censoring mechanism is random, the probability that an observation is censored. The ARL may be simulated or calculated approximately using the Markov chain approach, which we discuss in more detail in Chapter 11.

We propose using a parametric distribution such as the Weibull or log-logistic distribution to model the survival times primarily because the CUSUM is inherently a parametric procedure and it requires specifying the in-control and out-of-control values of a parameter that can be interpreted with respect to the quality of the process. While nonparametric estimates of the survival function are typically preferred in the medical literature, there are recent studies which employ parametric modeling of survival times [79, 80, 81, 82].
10.3.2 Modeling the survival distribution for each patient

Risk-adjustment for the RAST CUSUM chart is accomplished by incorporating information from the covariates into the RAST CUSUM score, which is defined by equation (10.13). There are a number of regression models for censored data that can be used to predict the survival time of a patient until time \( x \), given the value of the covariate vector \( u_i \). The popular Cox proportional hazards regression model uses a semi-parametric approach, where the baseline hazard rate (which is a function of the survival distribution) is modeled non-parametrically and a parametric form is assumed only for the effect of the covariate \( u_i \). However, the CUSUM chart is designed to detect changes in the process in terms of a change in a parameter of interest—and the semi-parametric form of the Cox model does not lend itself readily to this approach. In addition, the non-parametric estimate of the baseline hazard rate in the Cox model would only be defined on the range of the survival times that were present in the historical data set, thereby precluding extrapolation which might be necessary during prospective monitoring. For this reason, we use a parametric accelerated failure time (AFT) regression model to predict the survival time of each patient.

We note that the remainder of this section is based in part on the discussion of AFT regression models given by Klein and Moeschberger [83]. The AFT model is based on the assumption that that the survival function of patient \( i \) with covariate \( u_i \) at time \( x_i \) is the same as the baseline survival function at time \( x_i \exp\{\beta^T u_i\} \), i.e., \( S(x_i|u_i) = S_0(x_i \exp\{\beta^T u_i\}) \). The baseline survival function \( S_0 \) results when \( u_i = 0 \), and thus \( S_0(x_i) = S(x_i|u_i = 0) \). The term \( a_i = \exp\{\beta^T u_i\} \) is known as the acceleration factor. Increasing values of \( a_i \) correspond to increasing risk. For example, the probability of surviving at least until time \( x \) is higher for \( a_i = 0.5 \) then it is for \( a_i = 1 \). Figure 10.1 demonstrates the effect of the acceleration factor on the Weibull and log-logistic survival functions for various combinations of the shape parameter \( \alpha \) and the scale parameter \( \lambda \).

There are a number of parametric distributions that are traditionally used for AFT regression models. Among them are the Weibull (with the exponential as a special case), the log-normal, the log-logistic, and the generalized gamma distributions. The Weibull distribution is often used because of its flexibility and proven effectiveness in modeling failure times. In addition, the Weibull distribution is the only parametric distribution that can be formulated as both the AFT and the proportional hazards model. (Converting between the two models is accomplished by a simple re-parameterization). However, for purposes of comparing the RAST CUSUM chart to the RA Bernoulli CUSUM chart, we use the log-logistic distribution, since it is the only distribution whose AFT model can be represented as a proportional odds of failure model—a condition necessary to ensure the RAST CUSUM chart and the RA Bernoulli CUSUM chart are
Figure 10.1: Survival function of the Weibull and log-logistic distributions for different values of the acceleration factor, \( a \), and various combinations of shape parameter \( \alpha \) and scale parameter \( \lambda \).

designed to detect the same type of parametric shift. We will use the following parameterization of the log-logistic distribution:

\[
\begin{align*}
  f_0(x) &= \frac{\alpha}{\lambda} \left( \frac{x}{\lambda} \right)^{\alpha-1} \left[ 1 + \left( \frac{x}{\lambda} \right)^{\alpha} \right]^{-2} \\
  S_0(x) &= \left[ 1 + \left( \frac{x}{\lambda} \right)^{\alpha} \right]^{-1}
\end{align*}
\]  

where \( f_0(x) \) is the baseline density, \( S_0(x) \) is the baseline survival function, \( \alpha > 0 \) is the shape parameter and \( \lambda > 0 \) is the scale parameter. The expected value is \( \lambda \frac{\pi}{\sin(\pi/\alpha)} \) (if \( \alpha > 1 \)) and the distribution has support
on the positive real line. Using the AFT model for the log-logistic distribution, we now have

\[ S(x_i|u_i) = \left[ 1 + \left( \frac{x_i \exp(\beta^T u_i)}{\lambda} \right)^\alpha \right]^{-1}. \quad (10.15) \]

From equation (10.15) we can see that for patient \( i \), the scale parameter is now \( \lambda / \exp(\beta^T u_i) \) and thus the regression model is based on the assumption that \( X_i|u_i \sim \text{log-logistic}(\alpha, \lambda / \exp(\beta^T u_i)) \). Under baseline conditions, we have \( X_i|\{u_i = 0\} \sim \text{log-logistic}(\alpha, \lambda) \). In Section 10.3.3 we discuss how shifts in various measures of quality can be interpreted in terms of shifts in the parameter \( \lambda \).

For purposes of fitting the AFT regression model, it is convenient to write the natural logarithm of the survival time as a linear model. (This is sometimes called the log-linear representation of the AFT model). We begin by defining \( V \) as a random variable that follows the standard logistic distribution with density \( f_V(v) = \frac{e^v}{1 + e^v} \) and survival function \( S_V(v) = \frac{1}{1 + e^v} \). If we let \( \lambda = e^\mu \), \( \alpha = 1/\sigma \), and \( \beta = -\gamma \), we can readily demonstrate that for \( X|u \sim \text{log-logistic}(\alpha, \lambda / \exp(\beta^T u)) \), we have

\[ \log(X) = \mu + \gamma^T u + \sigma V. \quad (10.16) \]

We can then find the maximum likelihood estimates of \( \mu \), \( \sigma \), and \( \gamma \) by maximizing the likelihood function

\[ L(\mu, \sigma, \gamma|t_i, u_i) = \prod_{i=1}^n \left[ \frac{1}{\sigma} f_V \left( \frac{\log(t_i) - (\mu + \gamma^T u_i)}{\sigma} \right) \right] \delta_i \left[ S_V \left( \frac{\log(t_i) - (\mu + \gamma^T u_i)}{\sigma} \right) \right]^{1-\delta_i} \quad (10.17) \]

using historical observations indexed by \( i = 1, \ldots, n \). The log-linear representation is typically the one that is used by software packages, resulting in parameter estimates for \( \mu \), \( \gamma \), and \( \sigma \). The maximum likelihood estimates for \( \alpha \), \( \lambda \), and \( \beta \) can then be found using the invariance property of maximum likelihood estimators.

### 10.3.3 Specifying parameter shifts in the RAST CUSUM chart

We assume that the historical, in-control data set used to estimate \( \alpha \), \( \lambda \), and \( \beta \) is large enough to assume these parameters are known during the prospective monitoring of the patient survival times. In Section 10.3.1, we generically developed the RAST CUSUM scores to detect a change from \( \theta_i = \theta_{i0} \) to \( \theta_i = \theta_{i1} \). For both the Weibull and the log-logistic RAST CUSUM charts (which are discussed in more detail in Sections 10.4 and 10.5), we propose that \( \theta_{i0} = (\lambda_0 / \exp(\beta^T u_i), \alpha) \) and \( \theta_{i1} = (\rho_1 \lambda_0 / \exp(\beta^T u_i), \alpha) \), where \( \rho_1 \) is positive, but not equal to unity, and \( \lambda_0 \) represents the historical value of the baseline scale parameter \( \lambda \) when
the process was in control. Note that, in consequence of this definition, the shift from $\theta_{i0}$ to $\theta_{i1}$ presumes that $\alpha$ and $\beta$ remain unchanged, and that we are only interested in detecting a shift in the baseline scale parameter, $\lambda_0$.

We could, of course, define the shift from $\theta_{i0}$ to $\theta_{i1}$ so as to detect changes in the shape parameter $\alpha$ or even in $\beta$. However, for both the Weibull and log-logistic distributions, a shift in $\lambda$ is equivalent to the same shift in the mean survival time and the median survival time. Consequently, assuming that $\alpha$ and $\beta$ remain constant during prospective monitoring greatly simplifies the task of specifying meaningful values for $\theta_{i0}$ and $\theta_{i1}$, primarily because shifts in $\lambda$ are so easily interpretable. Assuming that $\beta$ is constant implies that the relationship between the covariate $U$ and the preoperative risk remains unchanged. This assumption is important, because if $\beta$ were to change, the legitimacy of the risk-adjustment is called into question. Hence, the adequacy of the AFT model fit should be assessed at regular intervals using recently acquired data during prospective monitoring. Likewise, the estimate of $\alpha$ should be assessed and updated if necessary at regular intervals. This, of course, presumes that the more recent historical data that are used to update the estimates of $\alpha$, $\beta$, and $\lambda_0$ correspond to periods of time when the process was in control.

In the preceding paragraph we mentioned that a shift in $\lambda$ is equivalent to the same shift in the mean survival time and the median survival time for the Weibull and log-logistic distributions. We demonstrate this below for the log-logistic distribution. The same results are easily shown for the Weibull distribution.

The mean survival time of patient $i$ for the log-logistic distribution is $MST_i = \frac{\lambda}{\exp(\beta^T u_i)} \frac{\pi}{\alpha \sin(\pi/\alpha)}$ (for $\alpha > 1$). Assuming that $\alpha$ and $\beta$ remain unchanged, detecting a shift from $MST_{i0} = \frac{\lambda_0}{\exp(\beta^T u_i)} \frac{\pi}{\alpha \sin(\pi/\alpha)}$ to $MST_{i1} = \rho_1 MST_{i0}$ is equivalent to detecting a shift from $\lambda_0$ to $\lambda_1 = \rho_1 \lambda_0$.

In addition, assuming a fixed $\alpha$ and $\beta$ allows us to discuss shifts in the median survival time in terms of $\lambda_0$. To see this, let $M_{i0}$ represent the in-control median survival time for patient $i$. Using equation (10.15) to solve $1 - S(M_{i0}|u_i) = 0.50$ for $M_{i0}$ we have

$$M_{i0} = \frac{\lambda_0}{\exp(\beta^T u_i)} \left( \frac{1}{1 - 0.5} - 1 \right)^{1/\alpha}. \quad (10.18)$$

Thus, detecting a shift from $M_{i0}$ to $M_{i1} = \rho_1 M_{i0}$ is equivalent to detecting a shift from $\lambda_0$ to $\lambda_1 = \rho_1 \lambda_0$. Furthermore, if we replace $M_{i0}$ in equation (10.18) with any quantile $q$ and insert the corresponding probability in place of 0.5, we can readily see that any multiple shift in $\lambda_0$ is equivalent to the same multiple shift in the quantile $q$, a property that holds for any family of distributions that is indexed only by a scale parameter (which is this case here, since we assume $\alpha$ and $\beta$ are fixed).
For the log-logistic model only, shifts in $\lambda$ are also related to shifts in the odds of dying before a fixed time, $c'$. This was the primary motivation for comparing the log-logistic RAST CUSUM chart to the RA Bernoulli CUSUM chart (as opposed to using the Weibull RAST CUSUM chart). If we assume that all survival times exceeding $c'$ are censored and that all observed $t_i < c'$ are not censored, then the odds of patient $i$ dying before time $c'$ are given by

$$\frac{1 - S(c' \mid u_i)}{S(c' \mid u_i)} = \left(\frac{c' \exp(\beta^T u_i)}{\lambda}\right)^\alpha.$$  \hspace{1cm} (10.19)

Using equation (10.19) we can represent a shift in the odds of death for patient $i$ in terms of a shift in the scale parameter $\lambda$:

$$\xi_1 = \frac{O_{11}}{O_{00}} = \left(\frac{c' \exp(\beta^T u_i)}{\lambda_1}\right)^\alpha \div \left(\frac{c' \exp(\beta^T u_i)}{\lambda_0}\right)^\alpha = \left(\frac{\lambda_0}{\lambda_1}\right)^\alpha = \rho_1^{-\alpha}. \hspace{1cm} (10.20)$$

Thus, the RAST CUSUM chart based on the log-logistic distribution and the RA Bernoulli CUSUM chart can both be calibrated to detect the same magnitude of shift in the odds of death, provided censoring occurs at a fixed value, $c'$.

### 10.4 The log-logistic RAST CUSUM chart

The purpose of this section is to define the log-logistic RAST CUSUM chart that will be compared to the RA Bernoulli CUSUM chart. Using the risk-adjusted survival function given by equation (10.15) and the corresponding density, we can build the likelihood ratio as defined by equations (10.12) and (10.13). For detecting a shift from $\lambda_0$ to $\lambda_1 = \rho_1 \lambda_0$, the score $W^L_i$ is given by

$$W^L_i(t_i, \delta_i) = \log \left(\frac{f(t_i, \rho_1 \lambda_0 \mid u_i) [S(t_i, \rho_1 \lambda_0 \mid u_i)]^{\delta_i}}{f(t_i, \lambda_0 \mid u_i) [S(t_i, \lambda_0 \mid u_i)]^{1-\delta_i}}\right). \hspace{1cm} (10.21)$$

which reduces to

$$W^L_i(t_i, \delta_i) = -\alpha \delta_i \log \rho_1 + 2\delta_i \left\{ \log \left[1 + \left(\frac{t_i \exp(\beta^T u_i)}{\lambda_0}\right)^\alpha\right] - \log \left[1 + \left(\frac{t_i \exp(\beta^T u_i)}{\rho_1 \lambda_0}\right)^\alpha\right] \right\}. \hspace{1cm} (10.22)$$
where the observed $t_i$ is defined in equation (10.11). Having defined the scores, the RAST CUSUM statistic is then calculated in the regular fashion as explained by equations (10.2) and (10.3) at the beginning of Section 10.2. The corresponding control limits for the log-logistic RAST CUSUM chart will be denoted by $h^+_L$ and $h^-_L$.

10.5 The Weibull RAST CUSUM chart

The adequacy of the parametric distribution used to model the survival times as well as the adequacy of the AFT regression model to describe the relationship of the risk covariates on patient survival are important considerations when implementing the RAST CUSUM chart. Because the Weibull distribution is one of the most popular parametric distributions for modeling survival times, and because the Weibull distribution is the only distribution that can be formulated as both the AFT model and as the well-known proportional hazards model, we present below a Weibull version of the RAST CUSUM chart.

We will use the following parameterization of the Weibull distribution:

$$
\begin{align*}
  f_0(x) & = \frac{\alpha}{\lambda} \left( \frac{x}{\lambda} \right)^{\alpha-1} \exp \left\{ -\left( \frac{x}{\lambda} \right)^\alpha \right\} \\
  S_0(x) & = \exp \left\{ -\left( \frac{x}{\lambda} \right)^\alpha \right\} 
\end{align*}
$$

(10.23)

where $f_0(x)$ is the baseline density, $S_0(x)$ is the baseline survival function, $\alpha > 0$ is the shape parameter and $\lambda > 0$ is the scale parameter. The expected value is $\lambda \Gamma(1 + 1/\alpha)$ and the distribution has support on the positive real line. Using the AFT model for the Weibull distribution, we now have

$$
S(x_i|u_i) = \exp \left\{ -\left( \frac{x_i \exp(\beta^T u_i)}{\lambda} \right)^\alpha \right\} .
$$

(10.24)

From equation (10.24) we can see that for patient $i$, the scale parameter is now $\lambda/\exp(\beta^T u_i)$. Hence the AFT regression model is based on the assumption that $X_i|u_i \sim \text{Weibull}(\alpha, \lambda/\exp(\beta^T u_i))$. Under baseline conditions, we have $X_i|\{u_i = 0\} \sim \text{Weibull}(\alpha, \lambda)$. As with the log-logistic model, shifts in the mean survival time and the median survival time are equivalent shifts in $\lambda$ when $\alpha$ and $\beta$ are held fixed. In addition, it is convenient to write the Weibull AFT model using the log-linear representation shown in equation (10.16), just as it was for the log-logistic AFT model. We can easily show that if $X \sim \text{Weibull}(\alpha, \lambda/\exp(\beta^T u))$ and if we let $\lambda = e^\mu$, $\alpha = 1/\sigma$, and $\beta = -\gamma$, then $V$ follows the extreme value distribution with density $f_V(v) = \exp(v - e^v)$ and survival function $S_V(v) = \exp(-e^v)$. We can then find the maximum likelihood
estimates of $\mu$, $\sigma$, and $\gamma$ by maximizing the likelihood function given by equation (10.17) using historical observations. The maximum likelihood estimates for $\alpha$, $\lambda$, and $\beta$ can then be found using the invariance property of maximum likelihood estimators.

From the Weibull AFT regression model, the survival function is given in equation (10.24). The corresponding risk-adjusted density is given by

\[
f(x_i|u_i) = \frac{\alpha \exp(\beta^T u_i)}{\lambda} \left( \frac{x \exp(\beta^T u_i)}{\lambda} \right)^{\alpha-1} \exp \left\{ - \left( \frac{x \exp(\beta^T u_i)}{\lambda} \right)^\alpha \right\}
\]

Using equations (10.24) and (10.25) to build the likelihood ratio in the manner shown by equation (10.21), the score for the Weibull RAST CUSUM chart for detecting a shift from $\lambda_0$ to $\lambda_1 = \rho_1 \lambda_0$ is given by

\[
W_i^W = (1 - \rho_1^{-\alpha}) \left( \frac{t_i \exp(\beta^T u_i)}{\lambda_0} \right)^\alpha - \delta_i \alpha \log \rho_1, \tag{10.26}
\]

where the superscript $W$ indicates the Weibull distribution and $t_i$ is defined in equation (10.11). Having defined the scores, the Weibull RAST CUSUM statistic would then be calculated as explained by equations (10.2) and (10.3) at the beginning of section 10.2.
Chapter 11

Determining the average run length

11.1 General considerations in using the ARL as a performance metric

We use the initial state average run length (ARL) to compare the RA Bernoulli CUSUM chart and the RAST CUSUM chart. Suitable estimates of $\lambda_0$, $\alpha$, and the regression parameter $\beta$ must be found using historical data. In addition, a specific shift size in the parameter of interest must be chosen. We denote these nominal shift sizes by $\xi_1$, the factor of change in the odds of failure, and $\rho_1$, the factor of change in the mean (or median) survival time. In contrast, $\xi$ or $\rho$ will represent any arbitrary shift that might occur, with $\xi = \rho = 1$ indicating no shift. The control limits $h^+$ and/or $h^-$ are chosen so that the in-control ARL is at least as large as a target value, $m_0$. Once the control limits are established, the ARL may be calculated for various values of $\xi$ or $\rho$ to assess the performance of the charts over a range of possible shifts. Because neither the RA Bernoulli CUSUM chart nor the RAST CUSUM chart have head-start features, a fair comparison between the charts can be made by comparing the initial state ARL performance of the charts. Using the initial state ARL presupposes that the shift either occurs prior to when monitoring begins or when the chart is at the zero state, which is furthest away from signal. Arguably, the steady-state ARL performance is the most realistic measurement of how we expect the charts to perform during prospective monitoring, since it only requires the assumption that the shift takes place sometime after the relatively rapid convergence to the steady-state distribution. (For an example of the rapid convergence to the steady-state distribution in the case of the unadjusted Bernoulli CUSUM chart, see Section 7.1.2). However, the RA Bernoulli CUSUM chart and the log-logistic RAST CUSUM chart are similar enough to assume that
the initial state $ARL$ performance should be adequate to make a fair comparison between the two methods, especially since calculation of the steady-state $ARL$ is more computationally intensive. We do not expect that using the initial state $ARL$ as the metric of comparison would lead to conclusions that differ from the conclusions we would have made using the steady-state $ARL$.

There are a number of approaches that can be used to calculate the $ARL$ of CUSUM charts [84]. Among them are simulation and the Markov chain approach. As the CUSUM statistic $Z_i$ is continuous, we must keep in mind that $ARL$ calculations arising from Markov chains are approximations because the Markov chain approach discretizes the continuous space in which $Z_i$ varies. In some situations, simulation of the $ARL$ is preferable—especially because it permits us to easily quantify the error of the $ARL$ estimate. A Monte Carlo replication of the RAST CUSUM score is obtained by first drawing an observation $u_i$ from the distribution of the covariate. If the censoring mechanism is random, $\delta_i$ would then be drawn from a specified Bernoulli distribution. Then $X_i$ or $C_i$ would be drawn from the survival or censoring distribution, conditional on $u_i$ and $\delta_i$, and then the simulated RAST CUSUM score could be calculated. A similar (but simpler) process can be followed to simulate the RA Bernoulli CUSUM chart.

Simulation may be necessary if calculation of the transition probabilities between the Markov states is sufficiently complicated or intractable. Suppose, for example, that censoring times in the RAST CUSUM chart are random. Then the $ARL$ of the chart depends on the distribution of four random variables: 1) the covariate, 2) the presence (or absence) of censoring, 3) the death times conditional on the covariate, and 4) the censor times conditional on the covariate. Accounting for each of these random variables when calculating the transition probabilities substantially increases the complexity of the problem. Furthermore, if the covariate is multidimensional, calculating the transition probabilities requires knowledge of the joint distribution of $U_i$.

In our comparisons of the RA Bernoulli CUSUM chart to the log-logistic RAST CUSUM chart that follow in Chapter 12, we used both the Markov chain approach and simulation to verify the estimates of the $ARL$. We simplified the problem by only using a single covariate, the Parsonnet score, and by assuming that for the log-logistic RAST CUSUM method, censoring only occurs at $c’ = 30$ days.

11.2 Markov chain approach for calculating the $ARL$

Brook and Evans [85] discussed using the Markov chain approach for calculating properties of the run length distribution of CUSUM charts where the chart statistic is continuous. The procedure involves
partitioning the space in which the CUSUM statistic varies into intervals, each interval corresponding to a state in the Markov chain. The transition probability matrix, $Q$, contains the transition probabilities $\pi_{jk} = P(Z_i \in \text{state } k \mid Z_{i-1} \in \text{state } j)$ for movement among the transient (non-absorbing) states. Note that $\pi_{jk}$ is the element in the $j^{th}$ row and $k^{th}$ column of the matrix $Q$. The average run length vector is given by $N = (I - Q)^{-1}1$, where $(I - Q)^{-1}$ is the fundamental matrix of the Markov chain and $1$ is a column vector of ones. Thus, for a Markov chain with $r$ states, the elements of the ARL vector $N^T = (N_1, N_2, \ldots, N_r)$ give the ARL if the chart were to begin in each of the corresponding states.

To calculate the transition probabilities, Brook and Evans originally proposed centering $Z_{i-1}$ in the middle of the corresponding interval in order to calculate the probability distribution of $Z_i$. To illustrate, let $(a, b]$ denote a state to which the CUSUM statistic will move, let $(c, d]$ indicate the current state of the CUSUM statistic, and let $m = (c + d)/2$ denote the midpoint of $(c, d]$. Brook and Evans used the approximation

$$P(a < Z_i \leq b \mid c < Z_{i-1} \leq d) \approx P(a < Z_i \leq b \mid Z_{i-1} = m) = P(a < Z_{i-1} + W_i \leq b \mid Z_{i-1} = m) = P(a - m < W_i \leq b - m) = F_W(b - m) - F_W(a - m)$$

(11.1)

where $F_W$ is the distribution function of the CUSUM score $W$. The accuracy of the approximation should improve as the number of states increases (and the width of the partition intervals decreases). Steiner et al. [56] advocated a rounding method approach that is similar to Brook and Evans’ centering method. Hawkins [86] proposed a refinement by assuming that, conditional on $c < Z_{i-1} \leq d$, the distribution of $Z_{i-1}$ is uniform. This gives

$$P(a < Z_i \leq b \mid c < Z_{i-1} \leq d) = P(a - Z_{i-1} < W_i \leq b - Z_{i-1} \mid c < Z_{i-1} \leq d) = \frac{1}{d-c} \int_c^d \{F_W(b - z) - F_W(a - z)\} \, dz$$

(11.2)

If $F_W$ is continuous, the integral in equation (11.2) is readily approximated using Simpson’s rule [40], where the interval $(c, d]$ is divided into two smaller intervals of equal size:

$$[\{F_W(b - c) + 4F_W(b - m) + F_W(b - d)\} - \{F_W(a - c) + 4F_W(a - m) + F_W(a - d)\}] / 6 .$$

(11.3)
Webster and Pettitt [87] demonstrated for the RA Bernoulli CUSUM chart that Hawkins’ smoothing approach performs more reliably than the rounding approach used by Steiner et al. [56]. That is, the smoothing method requires fewer partitions than the rounding approach to accurately approximate the $ARL$. Furthermore, as the number of partitions increases, the $ARL$ calculated using the smoothing method appears to converge monotonically to a constant value, whereas the convergence of the rounding method is slow and erratic. Webster and Pettitt [87] showed that even when the number of partitions is as high as 3500, the rounding method continues to fluctuate around the true $ARL$. For this reason, we use Hawkins’ smoothing method to calculate the transition probabilities.

In Sections 11.2.1 and 11.2.2, we outline the calculation of the Markov chain transition probabilities using Hawkins’ smoothing method for the RA Bernoulli CUSUM chart and the log-logistic RAST CUSUM charts. We describe the calculation of the transition probabilities for the upper CUSUM chart only. The approach is easily adapted to the lower CUSUM chart. However, additional care must be taken to calculate the $ARL$ of two-sided CUSUM chart schemes [88]. The transition probabilities of the log-logistic RAST CUSUM chart in Section 11.2.2 were derived under the simplifying assumption that all survival times exceeding a common value $c'$ are censored and that all observed $t_i < c'$ are not censored. Under the same censoring scheme, the transition probabilities for the Weibull RAST CUSUM chart could be readily derived by using the methodology that is shown for the log-logistic RAST CUSUM chart.

### 11.2.1 Transition probabilities for the RA Bernoulli CUSUM chart

The interval in which the CUSUM statistic can move without signaling, $[0, h_B^+]$, is partitioned into intervals of equal size such that each partition corresponds to a transient state in the Markov chain. Each transient state can be generically represented as an interval $(a, b]$, with the exception of the initial, or zero state, where $Z_i = 0$. The transition probabilities among the transient states depend on the distribution of the covariate $U$ and the distribution of $Y|U$, the binary outcomes conditional on the value of the covariate. The distribution of $U$ could be estimated from historical data using a parametric distribution or non-parametrically using the empirical cumulative distribution function. To begin with, we outline the approach for calculating the transition probabilities for a given value of the covariate, $u$. The transition probabilities can be categorized into four types: transitions 1) from a non-zero state to another non-zero state, 2) from a non-zero to the zero state, 3) from the zero state to a non-zero state, and 4) from the zero state to the zero state. We will use the subscript pairs (1, 1), (1, 0), (0, 1) and (0, 0) to represent transition probabilities in each of the four types, respectively.
To simplify matters, we begin by calculating the transition probabilities conditional on the value of the covariate and the outcome of the patient. Hence, we assume the value of \( u_i \) (and therefore \( p_i \)) as well as the value of \( y_i \) (and therefore \( w_i^B \)) are known. As stated above, we assume that, conditional on \( c < Z_{i-1} \leq d \), the distribution of \( Z_{i-1} \) is uniform on \((c, d]\). The conditional transition probability from a non-zero state to another non-zero state is given by

\[
\pi_{(1,1)}(w^B_i, u_i) = P(a < Z_i \leq b \mid c < Z_{i-1} \leq d) \\
= P(a - w_i^B < Z_{i-1} \leq b - w_i^B \mid c < Z_{i-1} \leq d) \\
= P(\max(a - w_i^B, c) < Z_{i-1} < \min(b - w_i^B, d)) / P(c < Z_{i-1} \leq d) \\
= \left[ \max(0, \min(b - w_i^B, d) - \max(a - w_i^B, c)) \right] / (d - c) .
\] (11.4)

For a transition from a non-zero state to a zero state, we have

\[
\pi_{(1,0)}(w^B_i, u_i) = P(Z_i = 0 \mid c < Z_{i-1} \leq d) \\
= P(Z_{i-1} \leq -w_i^B \mid c < Z_{i-1} \leq d) \\
= P(c < Z_{i-1} < \min(-w_i^B, d)) / P(c < Z_{i-1} \leq d) \\
= \left[ \max(0, \min(-w_i^B, d) - c) \right] / (d - c) .
\] (11.5)

The transition to the zero state from a non-zero state is given by

\[
\pi_{(0,1)}(w^B_i, u_i) = P(a < Z_i \leq b \mid Z_{i-1} = 0) \\
= I_{[a < w_i^B \leq b]} .
\] (11.6)

And the transition from the zero state to the zero state is given by

\[
\pi_{(0,0)}(w^B_i, u_i) = P(Z_i = 0 \mid Z_{i-1} = 0) \\
= I_{[w_i^B \leq 0]} .
\] (11.7)

To arrive at the unconditional transition probabilities, for each type of transition (i.e. for \( j = 0, 1 \) and \( k = 0, 1 \)) we must integrate over the values of \( y_i \) and \( u_i \) with respect to their measures,

\[
\pi_{(j,k)} = \int \left\{ P(Y_i = 1) \cdot \pi_{(j,k)}(w_i^B(1), u_i) + P(Y_i = 0) \cdot \pi_{(j,k)}(w_i^B(0), u_i) \right\} d\nu(u) \\
= \int \left\{ \frac{\xi p_i}{1 - p_i + \xi p_i} \pi_{(j,k)}(w_i^B(1), u_i) + \frac{1 - p_i}{1 - p_i + \xi p_i} \pi_{(j,k)}(w_i^B(0), u_i) \right\} d\nu(u)
\] (11.8)
where \( \nu(u) \) is the probability measure corresponding to the joint distribution of the covariate \( U \) and \( w^B_i(\cdot) \) is given by equation (10.10).

### 11.2.2 Transition probabilities for the log-logistic RAST CUSUM chart

In the derivation of the following transition probabilities, we assume all observations are censored (as necessary) at a common value \( c' \). As a result, we found it simplest to categorize the RAST CUSUM scores, \( W^L_i(t_i, \delta_i) \), into two groups: those that arise from the censoring (\( \delta_i = 0 \)) and those which arise from observed death times (\( \delta_i = 1 \)). Conditional that the death time of patient \( i \) is censored at time \( c' \), the four types of conditional transition probabilities discussed in Section 11.2.1 are identical to those given by equations (11.4), (11.5), (11.6), and (11.7), except replace \( w^B_i \) with \( w^L_i(c', 0) \) using equation (10.22). Further inspection of equation (10.22) reveals that for censored values, if \( \rho_1 < 1 \), then \( w^L_i(t_i, 0) < 0 \), which implies that \( \pi(0,1)|w^L_i(c', 0), u_i) = 0 \) and \( \pi(0,0)|w^L_i(c', 0), u_i) = 1 \).

If we condition on the event that a survival time is not censored, deriving the transition probabilities that are conditioned upon the value of the covariate require that we identify the cumulative distribution function of \( W^L_i \), conditional that the death time was actually observed, and not censored at \( c' \). We begin by defining the function \( g : W^L_i \to X \), which maps the log-logistic RAST CUSUM scores \( W^L_i \) to the observed death times, \( X \). Solving equation (10.22) for \( t_i \) when \( \delta_i = 1 \) gives:

\[
g(w^L) = \frac{\rho_1 \lambda_0}{\exp(\beta^T u)} \left[ \exp\left(\frac{1}{2}(w^L + \alpha \log \rho_1)\right) - 1 \right]^{\frac{1}{\rho_1}}. \tag{11.9}
\]

Note that when \( \rho_1 < 1 \), \( g \) is a decreasing function of \( w^L \) with domain \( (\alpha \log \rho_1, -\alpha \log \rho_1) \) and range \([0, \infty)\).

We now can compute the distribution function of \( W^L_i \):

\[
F_{W^L}(w) = P(W^L \leq w) = P(X > g(w)) = 1 + \left( \frac{g(w^L)}{\rho_1} \right) - \frac{1}{\rho_1} = 1 + \left( \frac{\rho_1}{\rho_1} \right) \left( \exp\left(\frac{1}{2}(w^L + \alpha \log \rho_1)\right) - 1 \right) \left( \frac{1}{\rho_1} - \exp\left(\frac{1}{2}(w^L + \alpha \log \rho_1)\right) \right)^{-1}. \tag{11.10}
\]

To find the distribution of \( W^L \) conditional that the observation was not censored (i.e. \( X \leq c' \)), we first
determine the probability of being censored:

\[ p_{cen} = P(X > c') = \left[ 1 + \left( \frac{c' \exp(\beta^T u)}{\rho \lambda_0} \right)^{\alpha} \right]^{-1}. \]  

(11.11)

In addition, we define the value of \( W^L \) when the outcome is censored:

\[ w_{cen} = W^L(c', 0) = \log \left[ 1 + \left( \frac{c' \exp(\beta^T u)}{\lambda_0} \right)^{\alpha} \right] - \log \left[ 1 + \left( \frac{c' \exp(\beta^T u)}{\rho_1 \lambda_0} \right)^{\alpha} \right]. \]  

(11.12)

Using equations (11.10), (11.11), and (11.12), we now have the \( F_{W^L|X \leq c'} \), the cumulative distribution of \( W^L \), given that the observation was not censored at \( c' \):

\[
F_{W^L|X \leq c'}(w^L) = \begin{cases} 
1 & \text{if } w^L > -\alpha \log \rho_1 \\
\frac{F_{W^L|(w^L-p_{cen})/p_{cen}} - p_{cen}}{1-p_{cen}} & \text{if } w_{cen} < w^L \leq -\alpha \log \rho_1 \\
0 & \text{if } w^L \leq w_{cen}.
\end{cases}
\]  

(11.13)

Determining the four types of transition probabilities conditional on the two events that the covariate is known and that the outcome is not censored is now straightforward. The conditional transition probability from a non-zero state to another non-zero state, \( \pi_{(1,1)}(w^L_i(x_i, 1), u_i) \) is given approximately by equation (11.3), except \( F_{W^L|X \leq c'} \) from equation (11.13) is used instead of \( F_W \). The conditional transition probability from a non-zero state to the zero state is given by

\[
\pi_{(1,0)}(w^L_i(x_i, 1), u_i) = P(Z_i = 0 \mid c < Z_i \leq d) = \frac{1}{\pi_c} \int_c^d F_{W^L|X \leq c'}(-z) \, dz 
\approx [F_{W^L|X \leq c'}(-d) + 4F_{W^L|X \leq c'}(-m) + F_{W^L|X \leq c'}(-c)] / 6.
\]  

(11.14)

The transition to a non-zero state from the zero state is given by

\[
\pi_{(0,1)}(w^L_i(x_i, 1), u_i) = P(a < Z_i \leq b \mid Z_i \leq d) = F_{W^L|X \leq c'}(b) - F_{W^L|X \leq c'}(a).
\]  

(11.15)
And the transition from the zero state to the zero state is given by

\[
\pi_{(0,0)}(w_i^L(x_i, 1), u_i) = P(Z_i = 0 \mid Z_{i-1} = 0) \\
= F_{W_i^L | X \leq c} (0) .
\]  

(11.16)

As with the RA Bernoulli CUSUM chart, to arrive at the unconditional transition probabilities, for each type of transition (i.e. for \( j = 0, 1 \) and \( k = 0, 1 \)) we integrate over the values of \( u_i \) for the two cases of censored and uncensored outcomes:

\[
\pi_{(j,k)} = \int \{ p_{cen} \cdot \pi_{(j,k)}(w_i^L(c', 0), u_i) + (1 - p_{cen}) \cdot \pi_{(j,k)}(w_i^L(x_i, 1), u_i) \} \, d\nu(u)
\]  

(11.17)

where \( \nu(u) \) is the probability measure corresponding to the joint distribution of the covariate \( U \).
Chapter 12

Comparisons

12.1 Cardiac surgery example

We motivate the comparison between the RA Bernoulli CUSUM chart and the log-logistic RAST CUSUM chart using the same cardiac surgery data that were used by Steiner et al. [56]. The data set describes 6,994 operations that were performed at a single surgical center in the United Kingdom over a seven-year period, 1992–1998. The data consist of information on each patient including operation date, surgeon, type of procedure, age, the Parsonnet score, and the number of days the patient lived after the operation (if death was observed during the seven year period). Like Steiner et al., we also selected the first two years as a training data set to estimate the in-control values of the chart parameters ($\lambda_0, \alpha$, and $\beta$). We assume that all recorded deaths in the training data set were related to the cardiac surgery. Survival times of patients who did not die during the study were censored at 30 days, as were the survival times of patients who died after 30 days. For both the RA Bernoulli CUSUM chart and the RAST CUSUM chart, we considered only a single covariate, the Parsonnet score, and we approximated its distribution $\nu(u)$ using the empirical distribution function calculated from the training data. The distribution of the actual Parsonnet scores in the training data set is shown in Figure 12.1.
12.2 Logistic regression versus the accelerated failure time regression model

To ascertain the influence of the Parsonnet score on the preoperative risk of failure, we assume logistic regression would be used to predict the preoperative risk for the RA Bernoulli CUSUM chart and the AFT regression model would be used for the RAST CUSUM chart. To compare how the two CUSUM charts would perform relative to one another in practice, we must also examine the impact of fitting the two different types of regression models used to predict preoperative risk. To this end, let us assume that the “true” stochastic process that gives rise to the patient outcomes is described by survival times which follow a log-logistic distribution. If a practitioner chooses to only observe and monitor 30-day mortality rates, the survival times are dichotomized into Bernoulli observations. If all survival times which exceed \( c' \) days are censored at \( c' \) days, then the deterministic components of the log-logistic AFT model and the logistic regression model are equivalent. To see this, we begin with the deterministic portion of the logistic regression model, where \( p \) denotes the probability of dying before time \( c' \):

\[
\log(\text{odds of dying before } c') = \text{logit}(p) = \beta_0 + \beta^T u .
\]

Figure 12.1: Histogram of the Parsonnet score for cardiac patients who underwent surgery in 1992 and 1993.
If we substitute the odds of dying before time $c'$ that is predicted by the AFT regression model (given by equation (10.19)) into equation (12.1) above, we have:

$$\logit(p) = \log \left( \frac{c' \exp(\beta^T u)}{\lambda} \right)^\alpha = \alpha \log (c'/\lambda) + \alpha \beta^T u.$$  \hspace{1cm} (12.2)

Hence $\alpha \log (c'/\lambda)$ corresponds to the intercept of the logistic regression model and the elements of $\alpha \beta$ correspond to the “slope” parameters of the logistic regression model. However, because estimates of the two models are obtained by maximizing different likelihood functions, the relationship demonstrated in equation (12.2) holds only approximately for a given set of data. For example, in the cardiac surgery data, fitting the AFT model to the first two years of data using the Parsonnet score as the single covariate and censoring all survival times (where necessary) at 30 days gives $\hat{\alpha} = 0.529$, $\hat{\lambda} = 30606.19$, and $\hat{\beta} = 0.1450$. (In the training data set, the 25 patients with survival times of 0 days were set to 0.5 days in order to make it possible to fit the AFT regression model). The relationship in equation (12.2) “predicts” that the logistic regression equation should be $\logit(p_i) = -3.6649 + 0.0767u_i$. This is very similar to the equation that results from actually fitting the logistic regression model to the same data: $\logit(p_i) = -3.6798 + 0.0768u_i$.

In the comparisons of the RA Bernoulli CUSUM chart and the RAST CUSUM chart that follow in Section 12.4, we only used the AFT regression models (instead of logistic regression) to predict the preoperative 30-day mortality rates. This was done to ensure that the comparison of the two charts would be as theoretically equitable as possible. However, in practice, if one only has binary data, then an AFT model could not be used to predict the 30-day mortality rate, and thus the RA Bernoulli CUSUM chart would necessarily be constructed using the predictions from logistic regression. However, we have demonstrated that the deterministic components of the two regression models are equivalent when a common censoring time is used, and that for the example cardiac data, the two regression models produce very similar results.

To further justify using only the AFT regression model to predict 30-day mortality when comparing the RA Bernoulli CUSUM and the log-logistic RAST CUSUM charts, we conducted a small simulation study to assess the difference between the AFT regression model and the logistic regression model in predicting 30-day mortality rates. To do this, we generated 150 data sets of $n = 2,218$ survival times, the number of observations in the training data set. In each simulated data set, the survival time for patient $i$ was randomly generated from the log-logistic($0.529, 30606.19/\exp(0.145 u_i)$) distribution where $u_i$ was the actual Parsonnet score for the $i^{th}$ patient in the training data set. The AFT model was then fit to the simulated survival times (which were censored as needed at 30 days), and the logistic regression was fit to the
Figure 12.2: Each curve represents the 30-day predicted mortality of the AFT model subtracted from the 30-day predicted mortality of the logistic regression model that were fit to one of the 150 simulated data sets. The difference in the predicted probability of 30-day mortality from the two fitted regression models was calculated across the range of observed Parsonnet scores (0 to 69). This process was repeated 150 times, and the differences in the prediction of 30-day mortality rates were calculated for each of the 150 simulated data sets. The results in Figure 12.2 demonstrate that the models differ the most (as much as ±8%) in predicting the 30-day mortality for those patients with larger Parsonnet scores (the higher risk patients). This is likely due to the sparseness of patients with high Parsonnet scores (see Figure 12.1). Although the two regression models do not produce identical results, they are, in our view, similar enough to justify using only the AFT regression model to predict 30-day mortality for the RA Bernoulli CUSUM chart for the purpose of comparing it to the RAST CUSUM chart.
12.3 Interpreting the accelerated failure time regression model

Examination of the AFT model estimates of $\hat{\alpha} = 0.529$, $\hat{\lambda}_0 = 30606.19$, and $\hat{\beta} = 0.1450$ from the training data reveals some interesting results. For example, this fitted model predicts that the median survival time after surgery for a low risk patient ($u = 0$) will be 30,606 days, or 83.8 years, which at first glance appears much too large. However, the AFT survival models should not be viewed as models that are designed to predict long-term survival, especially when age, arguably the most influential factor in human survival, is not included in the regression model. Rather, for our purposes, the AFT survival models should be viewed as models that reflect the proximal survival patterns that are most likely to be influenced by the quality of the surgery. One reason why the 30-day mortality rate is a reasonable measure of quality is because the causes of deaths that occur within 30 days of an operation are most likely to be related to the operation. Figure 12.3 shows how the non-parametric Kaplan-Meier [89] estimate of the survival function for the training data compares to the predicted survival function from the AFT model for Parsonnet scores of 0, 10, 20, and 30. The mean Parsonnet score in the training data for patients who survived past 30 days was 8.05 and was 21.3 for those who died within 30 days of the surgery. Keeping in mind that the Kaplan-Meier curve is not adjusted for the Parsonnet score, the shape of the survival curves predicted by the AFT regression model appears reasonable within the range of 0 to 30 days.

Figure 12.3: Visualizing the predicted survival function of the log-logistic AFT regression model for various values of the Parsonnet score ($u$) versus the non risk-adjusted Kaplan-Meier estimate of the survival function.
12.4 Comparisons of the RA Bernoulli CUSUM and the RAST CUSUM charts

Four cases were used to compare the performance of RAST CUSUM chart in terms of the performance of the RA Bernoulli CUSUM chart, each with a different expected fraction of censoring. For each case we assumed a known value of $\alpha = 0.529$ and that the target in-control $ARL$, $m_0$, was 10,000 patients. Given the frequency of the surgeries in this example, we can infer that $m_0 = 10,000$ corresponds to no more than one false alarm every nine years [56]. In addition, for each case, the charts were designed to be optimal at detecting a doubling in the odds of death, i.e., $\xi_1 = 2$, which corresponds to a decrease in the median time survival time by a factor of $\rho_1 = 0.2697$ (by equation (10.20)).

The training data have a 30-day mortality rate of 6.45%, which implies that about 93.5% of the observations are censored. In fact, the AFT regression model predicts that the expected number of baseline patients (Parsonnet score = 0) that live beyond 30 days to be higher than 97%, and thus the expected censoring fraction for baseline patients is also over 97%. Due to the high censoring fraction, the RAST CUSUM chart shows only minor improvement in the $ARL$ performance over the RA Bernoulli CUSUM chart. For this reason, we constructed three other cases by decreasing the baseline median survival time in order to reduce the fraction of censored observations and thereby demonstrate the performance gains of the RAST CUSUM chart as the censoring fraction decreases. We chose to reduce the median survival time in order to mimic clinical outcomes with high mortality rates. An alternative approach would have been to increase the censoring time $c'$ while holding the median survival time (i.e. $\lambda$) fixed. Changing the median survival time (i.e. changing the scale parameter $\lambda$) only stretches or contracts the survival distribution while maintaining the same shape. Consequently, for our purpose of investigating the $ARL$ performance of scenarios with lower censoring fractions, decreasing $\lambda$ will have the same impact on chart performance as increasing $c'$. In addition, increasing the censoring time is less appealing because the longer a patient lives past surgery, the less likely it is that the cause of death will be related to the surgery. Larger censoring times make it more difficult to assess the quality of the surgery in a timely fashion since patient outcomes must be plotted in the chart in the order they underwent surgery. However, the choice of the censoring time will also be influenced by clinical factors, and waiting a longer time to assess the outcomes of clinical procedures may, in fact, be advantageous. We discuss these ideas further in Section 13.2.

The parameter values used for the first case were those that were obtained from the training data set; the remaining three were synthetic, using the first case as a starting point. Specifically, for cases 2-4, values of $\lambda_0$ were first adjusted to achieve the target baseline censoring fraction of 75%, 50%, and 25% (when the
Parsonnet score was 0). Then, for each case, the value of \( \beta \) was adjusted so that the highest risk patient (with a Parsonnet score of 69) would have an expected censoring fraction that was 16% of the corresponding baseline censoring fraction. The same empirical distribution of the Parsonnet score that was calculated from the training data was used for all four cases. The parameter values for the four cases are shown in Table 12.1.

<table>
<thead>
<tr>
<th>Case</th>
<th>Censoring time</th>
<th>( m_0 )</th>
<th>( \xi_1 )</th>
<th>( \alpha )</th>
<th>( \lambda_0 )</th>
<th>( \beta )</th>
<th>Censoring fraction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Baseline</td>
</tr>
<tr>
<td>1</td>
<td>30 days</td>
<td>10,000</td>
<td>2</td>
<td>0.529</td>
<td>30.606</td>
<td>0.1450</td>
<td>0.975</td>
</tr>
<tr>
<td>2</td>
<td>30 days</td>
<td>10,000</td>
<td>2</td>
<td>0.529</td>
<td>239.360</td>
<td>0.0847</td>
<td>0.750</td>
</tr>
<tr>
<td>3</td>
<td>30 days</td>
<td>10,000</td>
<td>2</td>
<td>0.529</td>
<td>30.00</td>
<td>0.0669</td>
<td>0.500</td>
</tr>
<tr>
<td>4</td>
<td>30 days</td>
<td>10,000</td>
<td>2</td>
<td>0.529</td>
<td>3.76002</td>
<td>0.0569</td>
<td>0.250</td>
</tr>
</tbody>
</table>

Table 12.1: Parameter values for the four cases.

We acknowledge that the synthetic baseline censoring fractions of 75%, 50% and 25%, (which correspond to 30-day mortality rates of 25%, 50%, and 75%, respectively) represent significant (and even unrealistic) departures from the cardiac surgery data. However, we note that there do exist medical conditions whose clinical outcomes have high mortality rates within the first two years following the procedure. For example, the average three-month and one-year survival rates for a heart-lung transplant recipient are 70.5% and 58.2%, respectively [90]. In a recent review of randomized clinical trials for the treatment of high-risk melanoma, Verma et al. [91] reported two-year survival rates for patients in the treatment arms of these trials that ranged between 33% and 90%. Tagawa et al. [92] estimated the two-year survival of patients with “Stage IV melanoma who underwent surgical resection of metastatic sites followed by treatment on a peptide vaccine trial” to be around 60%. Larson et al. [93] reported a three-week overall survival rate for acetaminophen-induced liver failure of 71%. The censoring fractions for cases 2-4 were chosen primarily for the purpose of theoretically exploring the difference in the performance of the RA Bernoulli CUSUM and the RAST CUSUM charts and to roughly mimic procedures with high mortality rates.

For each of the four cases, the control limits that came as close as possible to the target value \( m_0 = 10,000 \) were identified for the log-logistic RAST CUSUM chart and the RA Bernoulli CUSUM chart. Adequate values for the control limits were readily found by relying on the fact that the natural logarithm of the in-control \( ARL \) is approximately a linear function of the control limit, \( h^+ \). The Markov chain approach (using Hawkins’ smoothing method) proved useful in locating the desired control limit. Once the control limits were identified, the initial state \( ARL \) was calculated for 21 different hypothetical shifts.
in the odds of 30-day mortality ($\xi = 1, 1.2, 1.4, \ldots, 4.8, 5$). Instead of using the Markov chain approach to approximate the $ARL$, at least 100,000 Monte Carlo replications were used to estimate the $ARL$ in order to be able to precisely quantify the estimation error. A total of $4 \times 21 = 84$ comparisons were made between the two CUSUM methods. A two-sided statistical test of hypothesis was conducted for each of the 84 comparisons, with the Bonferroni correction applied to maintain the overall Type I error rate at 5% across all 84 comparisons. Because the control limits were chosen to make the in-control $ARL$ of the two methods as close as possible, the statistical tests did not show significant differences between the $ARL$’s of the RA Bernoulli CUSUM and log-logistic RAST CUSUM charts for each of the four cases when $\xi = 1$. Among the 80 remaining tests that were conducted for $\xi > 1$, the $ARL$ of the log-logistic RAST CUSUM chart was always significantly less (all Bonferroni adjusted $p$-values were less than 0.0001) than the $ARL$ of the RA Bernoulli CUSUM chart, with only one exception: when $\xi = 1.2$ for case 1. The complete results of the four cases are shown in Tables B.1 and B.2 in Appendix B.

Figure 12.4: Ratio of the $ARL$ for the log-logistic RAST CUSUM chart to the $ARL$ of the RA Bernoulli CUSUM chart for the four cases, each with a different level of baseline censoring.

Figure 12.4 demonstrates the performance gains of the RAST CUSUM chart over the RA Bernoulli CUSUM chart in all four cases. The conclusion is obvious: the lower the censoring fraction, the more the RAST CUSUM chart outperforms the RA Bernoulli CUSUM chart. The decreasing curves in Figure 12.4 also indicate that larger shifts in the expected odds of death correspond to larger performance gains for
Figure 12.5: Average number of patients not exposed to higher risk as a result of using the log-logistic RAST CUSUM chart instead of the RA Bernoulli CUSUM chart. To give better resolution to case 1 (97.5% baseline censoring), the vertical axis is presented in the log scale.

The RAST CUSUM chart relative to the RA Bernoulli CUSUM chart. It is also evident that for very high levels of censoring, the performance of the two charts is practically the same. Assuming the frequency of the surgical procedure is relatively constant over time, if a doubling in the odds of expected 30-day mortality occurs, the RAST CUSUM chart signals an alarm only 1.03 times faster than the RA Bernoulli CUSUM chart if the baseline censoring fraction is 97.5%. For the smaller baseline censoring fractions of 75%, 50%, and 25%, the RAST CUSUM chart signals 1.12, 1.4, and 2.3 times faster than the RA Bernoulli CUSUM chart when a doubling in the expected odds of 30-day mortality occurs, respectively. The findings are intuitive, since fewer censored observations results in a larger fraction of observed death times, which in turn adds more information to the RAST CUSUM chart.

In addition to comparing the relative improvement in performance using the ratio of the ARL’s, we can also compare the CUSUM methodologies by examining how many patients, on average, would not be exposed to higher risk if a decrease in the quality of the procedure were to occur. In so doing, we assume that the amount of time needed to investigate and then improve the quality of the procedure after an alarm would be the same for both CUSUM methodologies. Figure 12.5 shows the average number of patients that would not be exposed to higher risk as a result of using the log-logistic RAST CUSUM chart instead of the RA Bernoulli CUSUM chart. Figure 12.5 demonstrates that the smaller the fraction of censored observations,
the larger the number of patients that will not be exposed to higher risk. Furthermore, the decreasing curves suggest that the smaller the increase in the odds of mortality, the larger the number of patients, who, on average, will not be exposed to the elevated risk conditions of the procedure. This highlights the increased sensitivity of the RAST CUSUM chart in detecting small decreases in procedural quality, especially when the censoring fraction is not too high.
Chapter 13

Discussion

13.1 Dynamics of the risk-adjusted CUSUM charts

To further illustrate how the risk adjustment works for the RA Bernoulli CUSUM chart and the log-logistic and Weibull RAST CUSUM charts, we examine the possible CUSUM scores that arise from the parameter estimates from the training data set of the cardiac surgery example. Examining the value of the score for various combinations of the covariate and patient outcomes makes it easier to understand how a risk-adjusted CUSUM statistic will react to each new observation.

For the RA Bernoulli CUSUM chart, the fitted logistic regression model for the training data set was $\logit(p_i) = -3.68 + 0.077u_i$, where $u_i$ is the Parsonnet score for patient $i$. Using equation (10.10), we examined how the score $W^B_i(y_i)$ changes for different values of the Parsonnet score and various patient outcomes when the nominal shift in the odds of failure is $\xi_1 = 2$. The numerical results are summarized in Table 13.1. The middle column of Table 13.1 demonstrates that the RA Bernoulli CUSUM statistic increases by a larger amount after the death of a low risk person than it does for a higher risk person. Hence, a higher penalty is assessed for the death of a low risk person. Conversely, the right-hand column shows that the CUSUM statistic decreases by a larger amount after the survival of a high risk person than than it does for the survival of a lower risk person. These results are logical and consistent with the behavior we would expect from a good risk-adjusted control chart. We also note that the for $\xi_1 > 1$, the RA Bernoulli CUSUM scores are bounded between $(-\log \xi_1)$ and $(\log \xi_1)$.

To better understand the dynamics of the RAST CUSUM scores given by equations (10.22) and (10.26), Figure 13.1 demonstrates how a patient’s survival time and their Parsonnet score ($u$) affect the
RAST CUSUM score. Each plot in Figure 13.1 consists of three curves that represent the RAST CUSUM score for three different values of the Parsonnet score when an actual death time is observed. Each plot also contains three solid circles plotted at day 30. These circles are positioned at the value of the RAST CUSUM score when the observation is censored at day 30 for each of the three example values of the Parsonnet score (0, 30, and 60). The parameter values used to create the Weibull RAST CUSUM chart were obtained by fitting the Weibull AFT model to the training data set of the cardiac surgery example. These estimated parameters were \( \hat{\alpha} = 0.4909 \), \( \hat{\lambda}_0 = 42133.6 \), and \( \hat{\beta} = 0.1307 \). The parameter estimates used for the log-logistic scores were the same as those discussed in Section 12.2. The scores for both types of RAST CUSUM charts were calculated using \( \rho_1 = 0.2697 \), which corresponds to \( \xi_1 = 2 \) for the log-logistic RAST CUSUM chart. Interestingly enough, the three censor scores for the log-logistic RAST CUSUM chart shown by the solid circles in Figure 13.1 are equivalent (to at least two decimal places) to the three survival scores for the RA Bernoulli CUSUM chart, which are shown in the right-hand column of Table 13.1. This result is probably attributable to the similarities in the log-logistic AFT and logistic regression models as discussed in Section 12.2.

We note that for both types of RAST CUSUM charts, when monitoring for decreases in the baseline mean survival time (\( \rho_1 < 1 \)), the score from a censored observations at time \( t \) is always less than the score for an observed death at time \( t \). This is precisely what we would expect, since a patient that is known to be alive at time \( t \) represents a higher level of quality than a patient who has died at the same time \( t \). The curves in Figure 13.1 clearly demonstrate that higher risk patients receive lower scores than lower risk patients, since the death of a high risk patient is less evidence of a degradation in quality than the death of a low risk patient. The curves also are decreasing, indicating that patients who live longer also provide less evidence of quality degradation than do patients with shorter survival times.

An important distinction between the Weibull and the log-logistic RAST CUSUM scores is that, when \( \rho_1 < 1 \), the log-logistic scores are bounded between \( (\alpha \log \rho_1) \) and \( (-\alpha \log \rho_1) \). However, the Weibull scores are only bounded above by \( (-\alpha \log \rho_1) \) and have no lower bound. This absence of a lower bound

<table>
<thead>
<tr>
<th>( u_i )</th>
<th>( W^B_i(1) ) (death)</th>
<th>( W^B_i(0) ) (survival)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.669</td>
<td>-0.024</td>
</tr>
<tr>
<td>30</td>
<td>0.509</td>
<td>-0.185</td>
</tr>
<tr>
<td>60</td>
<td>0.151</td>
<td>-0.542</td>
</tr>
</tbody>
</table>

Table 13.1: Illustration of the RA Bernoulli CUSUM score statistic, \( W^B_i(y_i) \), for various values of the Parsonnet score, \( u_i \).
in the Weibull score may be viewed as undesirable, because if a relatively large death time or censored time is observed for a high-risk patient, it has the potential to reset the CUSUM statistic \( Z_i^+ \) back to 0 even if the CUSUM chart had accumulated evidence of a degradation in quality and was close to signaling. Consequently, the log-logistic RAST CUSUM chart may be preferred as a monitoring tool over the Weibull RAST CUSUM chart, even if the Weibull model happened to fit the historical data slightly better than the log-logistic model.

It is interesting to note that the scores of the RA Bernoulli CUSUM chart and the log-logistic RAST CUSUM chart are both bounded between \((-\log \xi_1)\) and \((\log \xi_1)\) when \(\xi_1 > 1\) and the log-logistic RAST CUSUM chart is employed using the fixed censoring mechanism at time \(c'\) (since \(\xi_1 = \rho_1^{-\alpha}\) by equation (10.20)).

### 13.2 Prospective monitoring using RAST CUSUM charts

The best way to implement the RAST CUSUM chart as a prospective monitoring tool is not immediately obvious, since the order in which the death times or censored times are observed may not coincide with the order in which the operations were performed—especially if survival times longer than 30 days are considered. Plotting points in the RAST CUSUM chart in the order in which patient survival times arrive...
(as opposed to the order in which they receive treatment) could make it difficult to assess if and when the quality of a health care procedure has changed.

Suppose, for example, that the common censoring time was set to be one year after the operation. Again suppose we have been monitoring for several years and that a decrease in the quality of the surgeries occurs on January 1. Now some of the patients who are treated in January may die in February and their survival times would then be available in February. But patients treated in the previous year (when the process was in control) may have their survival times or censor times become available in, say, March. Thus, what would be observed in the months after January 1 could be a mixture of in-control and out-of-control observations. Consequently, this situation is not like the usual industrial situation where all observations which are taken after the special cause occurs are assumed to come from the out-of-control distribution. To avoid this problem, the outcome for each patient could not be included in the chart until one year had elapsed since the date of surgery, regardless of whether they had died earlier. Having to wait a year to get feedback on the process is likely to limit the usefulness of the chart as a tool for active quality improvement. In addition, deaths that occur long after a surgery are less likely to be related to the effects of the surgery, which would require more careful investigation of the cause of death so that the outcome could be correctly classified as a death related to the surgery or as a censored point.

All of this underscores the need to set a reasonable, and, if possible, relatively short censoring time at which patient outcomes are incorporated into the chart. This facilitates the inclusion of patient outcomes into the chart in the order they underwent the operation without waiting for extended periods of time—and plotting the patient outcomes in the order they underwent surgery permits us to distinguish between the in-control and out-of-control states. A short censoring time will also ensure that not too much time elapses before corrective action can be taken if the quality of the surgery changes. Axelrod et al. [66] aptly noted that “utilization of a short-term outcome is important if the CUSUM is to be successfully used as an active . . .management tool.” However, in some contexts, it may be necessary to use longer term outcomes (e.g. one year or eighteen months) in order observe clinical effects related to the procedure that are not likely to manifest themselves in a shorter period of time.

In the cardiac surgery example, we restricted the data by censoring all relevant death times at 30 days. Enforcing the common censoring time was done to provide a basis for comparing the RA Bernoulli CUSUM chart with the log-logistic RAST CUSUM chart. However, in practice, both the Weibull and the log-logistic RAST CUSUM charts can accommodate a variety of censoring patterns, because the form of the likelihood ratio in the RAST CUSUM score is invariant to the various types of right-censoring mechanisms.
we would likely encounter when monitoring clinical outcomes (see Section 10.3.1). Suppose we were to begin the prospective monitoring of a clinical procedure using a RAST CUSUM chart and that we decided that after 30 days, the survival time (or the censor time) of each patient would be incorporated into the chart in the order in which the patients underwent the procedure. There are four types of outcomes that could occur for each patient: 1) the patient dies of causes related to the procedure before 30 days, 2) the patient dies of causes unrelated to the procedure before 30 days, 3) we lose contact with the patient before 30 days has elapsed and we only know they were still alive at a certain time, or 4) the patient is still alive 30 days after the procedure. In this context, the first situation represents a true death time, and the remaining three situations describe observations that would be censored. Observing these different types of data over time can have some interesting consequences in a prospective monitoring scheme. We demonstrate these consequences with a hypothetical example.

Suppose that on May 10, we plot the RAST CUSUM chart, which will reflect the outcomes (censored and uncensored) of all surgeries performed up until April 10. Then on May 11, we add 4 more additional observations to the chart, in the order in which they received surgery on April 11. In addition, we also find out on May 11 that patient ‘A’, with whom we had lost contact 10 days after his surgery in February, has reappeared and is still alive. We also find out that patient ‘B’, who underwent the operation back in March, died on May 11. Originally, patient ‘A’ would have been included in the chart with a censored value of 10 days, and patient ‘B’ would have been included in the chart with a censoring time of 30 days. But we now know more information about patients ‘A’ and ‘B’, even though their 30-day outcomes have already been plotted in the RAST CUSUM chart.

This scenario begs the question of whether (and how) the censored observations for patients ‘A’ and ‘B’ should be updated at a later time when more information is available about their survival status. One approach would be to simply update the entire data set with new information about patient survival times and then rebuild the entire chart for a given period of monitoring, so that each patient is represented by only one point on the chart and the patients are plotted in the order in which they underwent the health procedure. Note that updating the chart with the new information from patients ‘A’ and ‘B’ presumes that the risk-adjustment model (i.e. the AFT regression model) reliably characterizes the survival time (or censor time) beyond 30 days—but risk-adjustment for survival or censor times which exceed 30 days may involve undesirable extrapolation. Furthermore, as we become aware of patients who die long after undergoing the operation, it will likely become increasingly difficult to determine whether the cause of mortality can be attributed to the quality of the surgical procedure.
Another consideration is that updating the RAST CUSUM chart would alter its run length properties—which could make it difficult to interpret the validity of an alarm because the in-control ARL would change. Nonetheless, even if the chart were not updated and all observations were censored at a common value $c'$, the results from Section 12.4 lead us to conclude that the RAST CUSUM methodology is still more sensitive to detecting increases in the mortality rate than the RA Bernoulli CUSUM chart. Furthermore, it stands to reason that even if the RAST CUSUM chart were not updated, if the censoring pattern were less restrictive than the common censoring value $c'$, that is, if we could observe all four types of patient outcomes mentioned previously, the RAST CUSUM chart would likely show even larger performance gains over the RA Bernoulli CUSUM chart because the more flexible censoring pattern would allow more information to be incorporated in the RAST CUSUM chart. For the sake of simplicity, perhaps it would be best not to update the chart at all. Yet, consciously excluding information regarding the quality of the process from the monitoring scheme is undesirable. We leave the investigation of how updating the chart would affect the $ARL$ properties to future research.

For historical assessments of the quality of a health care procedure, the issues surrounding prospective monitoring and updating become increasingly less of a concern the further back in time the operations were performed, since new information is less likely to surface about patients with censored survival times. The order in which the patient survival outcomes arrive versus the order in which patients underwent surgery is also a non-issue, since the vast majority (if not all) of the survival outcomes in a historical data set will already exist. In addition, longer censoring times pose less of a challenge when examining historical data.

### 13.3 Additional considerations

We have focused almost exclusively on upper CUSUM charts that are designed to detect mortality rates that are higher than those predicted by the risk-adjustment, corresponding to a decrease in procedural quality. The RAST CUSUM chart is more efficient than the RA Bernoulli CUSUM chart at detecting increases in the mortality rates, especially when the censoring fraction is not too high. The RAST CUSUM methodology can be easily applied to construct charts designed to monitor for improvements in surgical quality, as well. However, as mortality decreases, the censoring fraction is also likely to increase, depending on the nature of the censoring mechanism. And as the censoring fraction increases, the performance gap between the RAST CUSUM chart and the RA Bernoulli CUSUM chart narrows. Thus, the RA Bernoulli CUSUM chart may be preferred for situations when detecting improvements in procedural quality is of primary interest, especially
if the censoring fraction is expected to increase dramatically as a consequence of a decreasing shift in the patient mortality rates. However, the RAST CUSUM chart is not any more complicated to implement than the RA Bernoulli CUSUM chart, so any extra efficiency gained by using the RAST CUSUM chart would come at no additional cost.

The adequacy of the Weibull or log-logistic distributions as models of survival times and the adequacy of the AFT regression model in predicting the preoperative risk of failure must also be considered. Ideally, a large historical data set and careful model diagnostics would be used to ensure the validity of the models. The investigation of how the RAST CUSUM chart performs when the fit of the parametric survival models is less than ideal is a topic we recommend for future research. Furthermore, when estimating the parameters $\alpha$, $\lambda$, and $\beta$ from historical data, more information will typically be available about the patient than would usually be available during prospective monitoring. For example, in an effort to extract as much information as possible from the historical data, it may not be desirable to censor historical observations precisely at 30 days when fitting the AFT regression model, even though a 30-day censoring scheme would be used during prospective monitoring. This would probably allow the AFT model to more accurately predict the preoperative risk of patients whose censored CUSUM scores were updated at a later time. However, for a prospective monitoring scheme in which there was no updating, one could also argue that the AFT risk-adjustment model should be “trained” using the same type of data that would be incorporated into the chart during prospective monitoring—which, in this case, would mean that all historical observations should also be censored (as necessary) at 30 days prior to fitting the AFT model. In the cardiac surgery example that was used to compare the log-logistic RAST CUSUM and the RA Bernoulli CUSUM charts in Chapter 12, the AFT model was fit to historical observations that were censored (as necessary) at 30 days in order to make the comparison between the two CUSUM methodologies as equitable as possible.

Another topic for further research would be the investigation of how the RAST CUSUM chart performs under a variety of monitoring conditions. These conditions would be characterized by the value of the target in-control $ARL$, $m_0$, the size of the shift in the mean or median survival time that the chart is optimally designed to detect, $\rho_1$, the influence of risk-adjustment on the RAST CUSUM scores, and the distribution of the covariate(s).
Chapter 14

Conclusion

We set forth a new risk-adjusted CUSUM methodology for monitoring the survival times of patients after undergoing a health care procedure. The RAST CUSUM chart accounts for differences in the mortality rate for each patient by using the accelerated failure time regression model to predict the survival distribution of each patient while accommodating for a variety of censoring patterns. The log-logistic and the Weibull distributions were considered as possible parametric models of survival time that would be used to construct the RAST CUSUM chart. For detecting increases in the mortality rate, the log-logistic RAST CUSUM chart was shown to be more sensitive than the RA Bernoulli CUSUM chart, especially when the baseline censoring fraction is 75% or less. However, for detecting decreases in the patient mortality rates, we expect that the RAST CUSUM chart and the RA Bernoulli CUSUM chart would have very similar performance if the censoring fraction were to substantially increase as a result of a decrease in the mortality rate. The RAST CUSUM chart is an effective tool for monitoring increases in the mortality rate, especially if the censoring fraction is low, which would be the case for clinical outcomes that have high mortality rates. Its efficiency would be especially important for monitoring procedural outcomes that are associated with rare health conditions.

When implementing a prospective monitoring procedure using the RAST CUSUM chart, the order in which the patient outcomes are plotted in the chart must coincide with the order in which the patients underwent the surgery. When new information about censored outcomes that were previously included in the chart becomes available, it may be beneficial to update the chart by including that information in the data set and redrawing the chart. These issues are of less concern when assessing the historical quality of surgical outcomes in the relatively distant past. Special care should be taken to ensure that the Weibull or
log-logistic distributions provide reasonable models of the survival times, and that the accelerated failure
time regression model provides adequate prediction of the survival time of the patient based on one or more
covariates which influence the risk of mortality.

In addition to monitoring clinical outcomes, the RAST CUSUM chart could be used to monitor any
time-to-event variable whose survival distribution is influenced by one or more covariates. For example, it
could be used to detect increases or decreases in the useful life (or failure time) of a manufactured product.
In this context, risk-adjustment would mean accounting for various conditions of use (low use, high use,
harsh environment, etc.) as well as the various formulations in a family of similar products.
Bibliography


http://www.ustransplant.org/annual_reports/current/1311_h1.pdf [12 April 2006].


Appendix A

Graphical comparisons of Sets, SHDA, CUSCORE, and Bernoulli CUSUM methods
Case 1: $p_0 = 0.01$, $\gamma_1 = 1.5$, $m_0 = 150$

Case 2: $p_0 = 0.01$, $\gamma_1 = 4$, $m_0 = 900$

Case 3: $p_0 = 0.009$, $\gamma_1 = 3.75$, $m_0 = 100$

Case 4: $p_0 = 0.009$, $\gamma_1 = 4.75$, $m_0 = 200$

Case 5: $p_0 = 0.008$, $\gamma_1 = 2$, $m_0 = 50$

Case 6: $p_0 = 0.008$, $\gamma_1 = 2.25$, $m_0 = 550$

Figure A.1: SSANB profiles for cases 1 to 6.
Case 7: $p_0 = 0.008$, $\gamma_1 = 4.75$, $m_0 = 900$

Case 8: $p_0 = 0.007$, $\gamma_1 = 1.25$, $m_0 = 100$

Case 9: $p_0 = 0.007$, $\gamma_1 = 1.25$, $m_0 = 700$

Case 10: $p_0 = 0.007$, $\gamma_1 = 2.75$, $m_0 = 350$

Case 11: $p_0 = 0.007$, $\gamma_1 = 3.75$, $m_0 = 100$

Case 12: $p_0 = 0.006$, $\gamma_1 = 1.25$, $m_0 = 400$

Figure A.2: SSANB profiles for cases 7 to 12.
Figure A.3: SSANB profiles for cases 13 to 18.
Case 19: $p_0 = 0.004$, $\gamma_1 = 5$, $m_0 = 750$

Case 20: $p_0 = 0.004$, $\gamma_1 = 5.5$, $m_0 = 850$

Case 21: $p_0 = 0.004$, $\gamma_1 = 5.75$, $m_0 = 750$

Case 22: $p_0 = 0.003$, $\gamma_1 = 1.75$, $m_0 = 50$

Case 23: $p_0 = 0.003$, $\gamma_1 = 5.5$, $m_0 = 150$

Case 24: $p_0 = 0.002$, $\gamma_1 = 5.75$, $m_0 = 200$

Figure A.4: SSANB profiles for cases 19 to 24.
Case 25: $p_0 = 0.002$, $\gamma_1 = 6$, $m_0 = 800$

Case 26: $p_0 = 0.001$, $\gamma_1 = 3.25$, $m_0 = 100$

Case 27: $p_0 = 0.001$, $\gamma_1 = 6$, $m_0 = 150$

Case 28: $p_0 = 9e^{-04}$, $\gamma_1 = 3.5$, $m_0 = 850$

Case 29: $p_0 = 9e^{-04}$, $\gamma_1 = 4.5$, $m_0 = 50$

Case 30: $p_0 = 7e^{-04}$, $\gamma_1 = 4.5$, $m_0 = 350$

Figure A.5: $SSANB$ profiles for cases 25 to 30.
Figure A.6: SSANB profiles for cases 31 to 36. To reflect the variability in the simulated values of SSANB_{0}(\gamma) for the CUSUM, in cases 34 and 36, the thickness of the solid line for the CUSUM is equal to width of the 95% simultaneous Bonferroni confidence interval.
Case 37: $p_0 = 3e^{-04}$, $\gamma_1 = 2.25$, $m_0 = 600$

Case 38: $p_0 = 2e^{-04}$, $\gamma_1 = 4$, $m_0 = 750$

Case 39: $p_0 = 2e^{-04}$, $\gamma_1 = 5.25$, $m_0 = 100$

Case 40: $p_0 = 2e^{-04}$, $\gamma_1 = 5.75$, $m_0 = 350$

Figure A.7: SSANB profiles for cases 37 to 40. To reflect the variability in the simulated values of $SSANB_b(\gamma)$ for the CUSUM, in cases 37 to 40, the thickness of the solid line for the CUSUM is equal to width of the 95% simultaneous Bonferroni confidence interval.
Appendix B

Tabular comparisons of RA Bernoulli CUSUM and log-logistic RAST CUSUM charts

Tables B.1 and B.2 give the detailed results of the four cases used to compare the log-logistic RAST CUSUM chart and the RA Bernoulli CUSUM chart, indicated by the $L$ and $B$ subscripts, respectively. In the tables, $ARL$ represents the estimated initial state average run length evaluated at the corresponding hypothetical shift in the odds of death, $\xi$, and $SE$ represents the standard error of the corresponding $ARL$. All $ARL$ estimates in Tables B.1 and B.2 were based on at least 100,000 Monte Carlo replications.
<table>
<thead>
<tr>
<th>ξ</th>
<th>Case 1</th>
<th>Case 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ARL_L</td>
<td>ARL_B</td>
</tr>
<tr>
<td>1.0</td>
<td>10039.8</td>
<td>9976.5</td>
</tr>
<tr>
<td>1.2</td>
<td>2280.2</td>
<td>2295.8</td>
</tr>
<tr>
<td>1.4</td>
<td>859.9</td>
<td>876.3</td>
</tr>
<tr>
<td>1.6</td>
<td>456.0</td>
<td>468.2</td>
</tr>
<tr>
<td>1.8</td>
<td>298.8</td>
<td>306.4</td>
</tr>
<tr>
<td>2.0</td>
<td>220.3</td>
<td>227.1</td>
</tr>
<tr>
<td>2.2</td>
<td>175.0</td>
<td>179.9</td>
</tr>
<tr>
<td>2.4</td>
<td>145.3</td>
<td>149.8</td>
</tr>
<tr>
<td>2.6</td>
<td>124.6</td>
<td>128.2</td>
</tr>
<tr>
<td>2.8</td>
<td>109.5</td>
<td>113.3</td>
</tr>
<tr>
<td>3.0</td>
<td>97.9</td>
<td>101.1</td>
</tr>
<tr>
<td>3.2</td>
<td>88.8</td>
<td>91.7</td>
</tr>
<tr>
<td>3.4</td>
<td>81.4</td>
<td>84.1</td>
</tr>
<tr>
<td>3.6</td>
<td>75.1</td>
<td>77.6</td>
</tr>
<tr>
<td>3.8</td>
<td>70.1</td>
<td>72.4</td>
</tr>
<tr>
<td>4.0</td>
<td>65.7</td>
<td>67.9</td>
</tr>
<tr>
<td>4.2</td>
<td>61.8</td>
<td>64.0</td>
</tr>
<tr>
<td>4.4</td>
<td>58.6</td>
<td>60.5</td>
</tr>
<tr>
<td>4.6</td>
<td>55.6</td>
<td>57.5</td>
</tr>
<tr>
<td>4.8</td>
<td>53.1</td>
<td>54.9</td>
</tr>
<tr>
<td>5.0</td>
<td>50.7</td>
<td>52.6</td>
</tr>
</tbody>
</table>

Table B.1: Results from cases 1 and 2. For case 1, $h_L^+ = 4.832743$ and $h_B^+ = 4.798883$. For case 2, $h_L^+ = 6.0731$ and $h_B^+ = 6.006768$. 
Table B.2: Results from cases 3 and 4. For case 3, $h_L^+ = 6.241462$ and $h_B^+ = 6.021252$. For case 4, $h_L^+ = 6.27189$ and $h_B^+ = 5.5847$. 

<table>
<thead>
<tr>
<th>$\xi$</th>
<th>$ARL_L$</th>
<th>$ARL_B$</th>
<th>$SE_L$</th>
<th>$SE_B$</th>
<th>$ARL_L$</th>
<th>$ARL_B$</th>
<th>$SE_L$</th>
<th>$SE_B$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>10041.4</td>
<td>9999.1</td>
<td>31.5721</td>
<td>14.0512</td>
<td>10010.0</td>
<td>10008.8</td>
<td>14.0757</td>
<td>14.0271</td>
</tr>
<tr>
<td>1.2</td>
<td>1138.2</td>
<td>1264.4</td>
<td>3.4332</td>
<td>3.7798</td>
<td>1117.5</td>
<td>1520.0</td>
<td>3.3165</td>
<td>6.1119</td>
</tr>
<tr>
<td>1.4</td>
<td>312.3</td>
<td>383.9</td>
<td>0.8226</td>
<td>0.9857</td>
<td>302.8</td>
<td>527.9</td>
<td>0.7984</td>
<td>1.3239</td>
</tr>
<tr>
<td>1.6</td>
<td>98.0</td>
<td>134.7</td>
<td>0.1807</td>
<td>0.2419</td>
<td>95.0</td>
<td>209.4</td>
<td>0.1733</td>
<td>0.3698</td>
</tr>
<tr>
<td>1.8</td>
<td>73.9</td>
<td>104.1</td>
<td>0.1201</td>
<td>0.1627</td>
<td>71.7</td>
<td>166.3</td>
<td>0.1147</td>
<td>0.2575</td>
</tr>
<tr>
<td>2.0</td>
<td>59.9</td>
<td>86.8</td>
<td>0.0873</td>
<td>0.1216</td>
<td>58.5</td>
<td>141.3</td>
<td>0.0843</td>
<td>0.1946</td>
</tr>
<tr>
<td>2.2</td>
<td>51.1</td>
<td>75.5</td>
<td>0.0688</td>
<td>0.0970</td>
<td>50.1</td>
<td>125.4</td>
<td>0.0666</td>
<td>0.1577</td>
</tr>
<tr>
<td>2.4</td>
<td>45.0</td>
<td>67.7</td>
<td>0.0565</td>
<td>0.0811</td>
<td>44.3</td>
<td>114.2</td>
<td>0.0548</td>
<td>0.1331</td>
</tr>
<tr>
<td>2.6</td>
<td>40.5</td>
<td>62.2</td>
<td>0.0479</td>
<td>0.0696</td>
<td>39.9</td>
<td>105.7</td>
<td>0.0465</td>
<td>0.1148</td>
</tr>
<tr>
<td>2.8</td>
<td>37.3</td>
<td>57.9</td>
<td>0.0421</td>
<td>0.0614</td>
<td>36.6</td>
<td>99.4</td>
<td>0.0407</td>
<td>0.1015</td>
</tr>
<tr>
<td>3.0</td>
<td>34.5</td>
<td>54.3</td>
<td>0.0373</td>
<td>0.0547</td>
<td>34.0</td>
<td>94.4</td>
<td>0.0362</td>
<td>0.0914</td>
</tr>
<tr>
<td>3.2</td>
<td>32.4</td>
<td>51.7</td>
<td>0.0336</td>
<td>0.0500</td>
<td>32.0</td>
<td>90.3</td>
<td>0.0327</td>
<td>0.0837</td>
</tr>
<tr>
<td>3.4</td>
<td>30.6</td>
<td>49.4</td>
<td>0.0307</td>
<td>0.0455</td>
<td>30.2</td>
<td>86.8</td>
<td>0.0299</td>
<td>0.0770</td>
</tr>
<tr>
<td>3.6</td>
<td>29.0</td>
<td>47.4</td>
<td>0.0283</td>
<td>0.0425</td>
<td>28.7</td>
<td>84.0</td>
<td>0.0274</td>
<td>0.0715</td>
</tr>
<tr>
<td>3.8</td>
<td>27.8</td>
<td>45.8</td>
<td>0.0261</td>
<td>0.0397</td>
<td>27.5</td>
<td>81.6</td>
<td>0.0254</td>
<td>0.0667</td>
</tr>
<tr>
<td>4.0</td>
<td>26.6</td>
<td>44.5</td>
<td>0.0244</td>
<td>0.0372</td>
<td>26.4</td>
<td>79.5</td>
<td>0.0239</td>
<td>0.0627</td>
</tr>
<tr>
<td>4.2</td>
<td>25.6</td>
<td>43.2</td>
<td>0.0229</td>
<td>0.0351</td>
<td>25.5</td>
<td>77.8</td>
<td>0.0225</td>
<td>0.0602</td>
</tr>
<tr>
<td>4.4</td>
<td>24.8</td>
<td>42.2</td>
<td>0.0217</td>
<td>0.0334</td>
<td>24.6</td>
<td>76.1</td>
<td>0.0211</td>
<td>0.0566</td>
</tr>
<tr>
<td>4.6</td>
<td>24.0</td>
<td>41.2</td>
<td>0.0206</td>
<td>0.0317</td>
<td>23.8</td>
<td>74.8</td>
<td>0.0200</td>
<td>0.0542</td>
</tr>
<tr>
<td>5.0</td>
<td>23.4</td>
<td>40.3</td>
<td>0.0196</td>
<td>0.0303</td>
<td>23.2</td>
<td>73.5</td>
<td>0.0191</td>
<td>0.0517</td>
</tr>
</tbody>
</table>