Performance Comparison of Statistical Process Control Methods to Monitor Grouped Data and Rare Events in Healthcare Applications

A Thesis Presented

by

Zeynep Karakus

to

The Department of Industrial Engineering

in partial fulfillment of the requirements

for the degree of

Master of Science

in

Industrial Engineering

Northeastern University
Boston, Massachusetts
(April 2013)
Abstract

Health expenditures in the United States reached $2.6 trillion in 2010, over ten times the $256 billion spent in 1980. In 2011, national health spending is estimated to have neared $2.7 trillion. While there is broad agreement that the increase in costs must be controlled, there is disagreement over the driving issues. To improve healthcare performance, we must change our way of working; however, change does not always indicate improvement. The changes that produce improvement and those that do not need to be distinguished and related aspects of performance must be measured. Health care costs are an ever growing burden and we seek to help by engineering systems that preserve quality while improving the system. Statistical Process Control (SPC) is a statistical method that may facilitate decision-making about where improvement efforts should be focused. Control charts, central to SPC, are used to monitor and analyze the performance and variation of a process, including biological processes such as blood pressure homoeostasis or organizational processes such as patient waiting times in a hospital, over time. Variation affects almost every performance measure of the Institute of Medicine's six key dimensions of a good health care system: efficiency, effectiveness, safety, satisfaction, access and equity. There are three different variation types here; variability in processes, practices and people. Although SPC methods are widely used in manufacturing to monitor and improve process variability, its importance is not acknowledged well in health care. Investigators can identify process performance that is not satisfactory, and therefore in need of refinement, using SPC applications which also have a helpful role in determining the effects of interventions made to processes. This study consists of theoretical and practical approaches in order to find the best way for monitoring health care processes as well as applications of SPC in healthcare systems such as Neonatal Quality Improvement Collaborative of Massachusetts (NeoQIC) and New England Veterans Affairs (VA).
Acknowledgements

I would like to express the deepest appreciation to my adviser and committee chair, Dr. James C. Benneyan, who supported me with his sound statistics background and tremendous scholarship. Thanks to him, I have been able to make my three desires real; education abroad, working in projects related to statistics, and working in health care. In addition, I consider it an honor to work with New England VERC who has provided scholarship for my education.

I would like to express my sincere gratitude to my committee members, Dr. Thomas Cullinane and Dr. Dayna Martinez, also thanks to all my colleagues for reviewing the papers. I am indebted to my all colleagues and friends in QPL who supported me, especially to Hande Musdal, Onder Ondemir, Serpil Mutlu and Oguz Erdogan. Their support ensured my determination and patience during last three years.

I would like to thank my loved ones, who have supported me throughout entire process, both by keeping me harmonious and helping me to shape my today and tomorrow. This thesis is dedicated to my parents who have presented me the opportunity of an education from the best institutions and help throughout my life. This thesis would not happen to be possible without my parents Ayfer and Fahri Karakus, my sister Yagmur Karakus, and the person who is always with me; Kaan Ozturk. I will be grateful forever for your love.
Table of Contents

Abstract .............................................................................................................................................................. ii
Acknowledgements ........................................................................................................................................ iii
List of Figures ................................................................................................................................................ vii
List of Tables ................................................................................................................................................... ix
Introduction .................................................................................................................................................... 1
  Why Health Care and IE? .............................................................................................................................. 1
  Why SPC Methods? .................................................................................................................................. 3
History of SPC and Previous Applications in Health Care ........................................................................ 4
Brief Description of the Thesis ....................................................................................................................... 7
Chapter 1. Applications of SPC to Neonatology Quality Improvement ......................................................... 9
  1. Background ........................................................................................................................................ 9
  2. Methodology ..................................................................................................................................... 10
  3. Results ............................................................................................................................................ 12
    3.1. \( p \) chart for infection rate ........................................................................................................... 12
    3.2. \( g \) chart for the number between infections ............................................................................. 13
    3.3. \( \bar{X} \) and \( S \) control charts for temperature ........................................................................... 13
    3.4. XMR, EWMA and CUSUM charts for temperature ..................................................................... 14
    3.5. The relationship between temperature and infection variables ................................................ 15
    3.6. \( p \) chart for \( T^* \) ............................................................................................................................ 16
    3.7. Relative performances of the charts ............................................................................................ 17
  4. Discussion .......................................................................................................................................... 17
Chapter 2. Relative Performance of Shewhart and EWMA \( p \) and \( g \) Statistical Control Charts for Monitoring Adverse Healthcare Event Rates ......................................................... 19
  1. Background ....................................................................................................................................... 19
  2. Methodology ..................................................................................................................................... 20
2.1. $p$, $g$, and EWMA $g$ charts ................................................................. 20

2.2. Choosing the parameters of EWMA chart ............................................. 23

2.3. Performance Analysis ............................................................................ 24

3. Results ......................................................................................................... 26

3.1. Shewhart $p$ vs $g$ chart for low rates ..................................................... 26

3.2. Improving $g$ charts ............................................................................... 27

4. Discussion ................................................................................................... 31

Chapter 3. Development of an Online Improvement Dashboard for Specialty Care in Veterans Affairs .. 32

1. Background ............................................................................................... 32

2. Content of the Compass ............................................................................ 35

3. Quality Measures ....................................................................................... 36

3.1. Access measures .................................................................................... 36

3.2. Care coordination measures ................................................................. 38

3.3. FOBT tracking ...................................................................................... 38

3.4. DAPTposDES tracking ......................................................................... 39

4. Monitoring Methodologies ........................................................................ 40

4.1. Run chart .............................................................................................. 40

4.2. $p$ chart ............................................................................................... 40

4.3. Individuals chart .................................................................................. 41

5. Sample Outputs from a VA Facility .......................................................... 42

6. Discussion .................................................................................................. 47

Chapter 4. Statistical Process Control Methods for Grouped Data in Normal and Weibull Models ...... 48

1. Background ............................................................................................... 48

2. Methodologies for Monitoring Grouped Data .......................................... 51

2.1. The midpoint approach – Quantile chart .............................................. 51

2.2. Maximum likelihood estimation ............................................................ 54

2.3. $p$ chart ............................................................................................... 56
2.4. Chi-square control chart ................................................................. 56
2.5. The method of minimum chi-square ............................................. 57
2.6. Simulation with theoretical approaches: MLE and minimum chi-square method ................. 58
3. Simulation Studies ........................................................................... 59
  3.1. Detecting the shift in the process mean ........................................ 61
  3.2. Detecting the shift in the process standard deviation ..................... 65
4. Application to the VA Colonoscopy Services .................................... 67
5. Discussion ......................................................................................... 70
Conclusion ........................................................................................... 72
References ............................................................................................ 75
Appendix ............................................................................................... 81
List of Figures

Figure 1. Percent of patients in Massachusetts that did not get or receive delayed medical care due to cost ................................................................. 1

Figure 2. The number of published articles on SPC in health care. Black bars: studies conducted in the USA; grey bars: studies outside the USA ................................................................. 5

Figure 3. General format of a control chart ................................................................. 6

Figure 4. (a) $p$ chart and (b) standardized $p$ chart for infection data ................................................................. 13

Figure 5. $g$ control chart for infection data ................................................................. 13

Figure 6. (a) S, (b) $\bar{X}$ with S charts for the temperature variable ................................................................. 14

Figure 7. (a) Individual chart, (b) EWMA, (c) MR, and (d) CUSUM control charts for the temperature variable ................................................................................................. 15

Figure 8. (a) $p$ chart and (b) standardized $p$ chart of $T^*$ with variable subgroup sample sizes ................................................................. 16

Figure 9. (a) OC curves and (b) ARL curves for $\bar{X}$ chart and $p$ chart of temperature ................................................................. 17

Figure 10. (a) Probability in-control of $k = 3$ sigma $p$ and $g$ charts for positive rate changes, ................................................................. 26

Figure 11. Probabilities of being in-control for $n = 1$ when (a) $p = 0.1$ and (b) $p = 0.00001$ ................................................................. 28

Figure 12. Probabilities of being in-control for $n = 10$ when (a) $p = 0.1$ and (b) $p = 0.00001$ ................................................................. 30

Figure 13. IHI’s Collaborative model for achieving Breakthrough improvement ................................................................................................. 34

Figure 14. CRC screening process ................................................................................................. 39

Figure 15. Screenshot of data input tab for national data followed by facility data ................................................................................................. 42

Figure 16. Core appointment access measures; average days until 3rd next available appointment, e-consults and average new patient wait time ................................................................................................. 42

Figure 17. Support access measures; appointments scheduled within 14 days and missed opportunities ................................................................................................. 43

Figure 18. Percent of missed opportunities instead of plotting numbers in Figure 17 ................................................................................................. 44

Figure 19. Care coordination measures; octane ratio and incomplete consults ................................................................................................. 45

Figure 20. FOBT tracking tab ................................................................................................. 46
Figure 21. DAPTpostDES measures ........................................................................................................... 46

Figure 22. (a) $\bar{X}$ chart with midpoint approach, (b) data in the form of a frequency distribution .......... 52

Figure 23. Flow chart of Monte Carlo simulation with Newton Raphson method to estimate $\lambda$ of Weibull distribution .......................................................................................................................... 59

Figure 24. Flow chart of Monte Carlo simulation to monitor process mean .............................................. 60

Figure 25. ARL curves for case 1 ............................................................................................................... 61

Figure 26. ARL curves for case 4 ............................................................................................................... 62

Figure 27. ARL values for different interval widths ($0.2\sigma, 0.4\sigma, 0.6\sigma, 0.8\sigma, 0.9\sigma, 1.2\sigma$ are $w = 5, 10, 15, 20, 24, 30$ respectively) when (a) the process is in-control and (b) the shift parameter is $\gamma = 0.1$ ............ 65

Figure 28. Bias in the mean and standard deviation estimations respectively by midpoint method......... 67

Figure 29. (a) Complete dataset of days to colonoscopy after positive FOBT, (b) Cumulative run chart for grouped data used by the VA ...................................................................................................................... 68

Figure 30. Monitoring (a) mean and (b) standard deviation with classic and midpoint approach............. 69

Figure 31. (a) Chi-square chart for multivariate monitoring, (b) $p$ chart for proportion of days $<60$ .......... 70
List of Tables

Table 1. Variables and their definitions ........................................................................................................ 10
Table 2. Binary logistic regression for the infection and temperature variables ........................................ 16
Table 3. Control limits for $g$ charts ........................................................................................................... 22
Table 4. Control limits for EWMA chart .................................................................................................... 23
Table 5. ARLs of EWMA chart for different $\lambda$ and $k$ values for $p = 0.01$ ........................................ 24
Table 6. Average Run Lengths for EWMA and $g$ charts with different scenarios for $n = 1$ .................. 28
Table 7. ARLs for optimal sample size $n = 10$ ...................................................................................... 29
Table 8. ARLs when in-control ARLs are common (0 shift represents in-control) ................................. 30
Table 9. Dashboard measures aligned to the Triple Aim and Six Dimensions of IOM ............................ 34
Table 10. Definitions of the measures included in the dashboard ............................................................ 35
Table 11. Control limits for individual observations and moving range (MR) .......................................... 41
Table 12. An example for grouped data; patient appointment waits in a facility ................................... 50
Table 13. Grouped data with $\bar{x} = 69.691$ and $s = 17.157$ by midpoint approach .............................. 52
Table 14. General formulation of grouped data ....................................................................................... 54
Table 15. Control limits for chi-square chart ........................................................................................... 57
Table 16. Estimating $\lambda$ with Excel Solver .......................................................................................... 59
Table 17. Average run lengths for each method with different settings to detect shifts in the process mean ........................................................................................................................................................... 63
Table 18. ARLs for each method with different settings to detect shift in Normal process variance ........ 66
Table 19. Sample data set from colonoscopy services of the VA ............................................................. 68
Table 20. Frequencies of colonoscopy in 30, 60, 90, 120, >120 days, and sample mean and standard deviations estimated by classic and midpoint approach ................................................................. 69
Introduction

Why Health Care and IE?

Health care is one of the top social and economic concerns in the United States today. The increasing cost of medical care and health insurance is affecting the quality of life for many Americans in one way or another. From lack of access to preventative care and the unaffordability of medical treatment, Americans are facing many health care problems. Not only are the uninsured people unable to pay the necessary medical care, it is becoming difficult for those with health insurance as well. The United States has one of the highest rates for health care costs. Health care expenditures in the United States are at 15.3% of GDP which is the highest of any developed country [1]. An excess of money is being spent on health care but it is not being used efficiently. An estimated one-third of 2006 health care expenditures, which is nearly 5% of GDP, did not fix health care systems. The National Health Interview Survey conducted in 2011, for example, points out that the percent of patients who did not get or receive delayed medical care or did not get pharmaceuticals due to cost has gradually increased over the years. Figure 1 shows the percent that did not get or receive medical care due to cost in Massachusetts.

Several studies agree that about 30% of total U.S. healthcare costs are due to ineffective, poorly designed processes, prompting publications by the National Academy of Engineering (NAE) and Institute of Medicine (IOM) to promote increased application of systems engineering, operations research, management science, and related methods used in other industries [2]. Each IOM dimensions of the care system - efficiency, effectiveness, safety, access, equity, patient-centered - can be improved by systems
engineering, including patient flow, appointment and staff scheduling, facility layout, disease screening, treatment optimization, regional planning, and pandemic response problems. Systems engineers use several methods in order to model, analyze, predict, improve, and optimize the performance of complex systems. Although these methods are widely used within clinical, operations, and administrative healthcare processes, their focus has been insignificantly positioned to dramatically move the U.S. healthcare system forward. As emphasized in a recent article by Pronovost, “It is time for the science of health care delivery to mature and embrace systems engineering” [3]. Measuring performance is an essential step in guiding quality improvement. Quality improvement techniques in health care can help provide a foundation for standardization and systematic improvement that can facilitate lower preventable and avoidable harm. The global health care community has widely divergent perspectives on the use of data and information for decision-making. In order to provide patients with the best possible care, an assessment of the current state will help guide improvement planning.

This research is funded by the New England Veterans Engineering Resource Center (NE VERC) which is one of four national Veterans Engineering Research Centers created in July 2009 that together represent the single largest investment in healthcare process improvement in the U.S. NE VERC partners with Northeastern University’s healthcare industrial engineering program, the Massachusetts Institute of Technology, Worcester Polytechnic Institute, and several Veterans Affairs (VA) centers of excellence. The partnership develops and implements industrial engineering solutions to improve process quality, access, safety, efficiency, and performance throughout the VA healthcare system, as well as New England and U.S. healthcare processes.
Why SPC Methods?

A vast amount of data is collected routinely from various healthcare processes. The analysis of this data can provide understanding into the behavior of these healthcare processes. Statistical Process Control (SPC) methods can be used to highlight areas that need further investigation as well as identify variation within the process. Exploring this process variation is the first step for quality improvement. It guides to the type of action that is proper to improve the functioning of a process. Variation is everywhere and it is unavoidable. Achieving steady quality requires understanding, monitoring and controlling the variation. A continuous commitment is necessary in order to obtain optimal product quality.

Montgomery defines SPC as “a powerful collection of problem-solving tools useful achieving process stability and improving capability through the reduction of variability” [4]. There are various SPC techniques that can be applied to data. Two main tools of SPC are control charts and process capability analysis. Additional tools useful in process improvement include cause-and effect diagrams, flow charts, Pareto diagrams, and check sheets. The purpose of these methods is to identify when the process is exhibiting unusual behavior. SPC methods are useful in highlighting these unusual activities; however, these methods do not necessarily show that the process is either right or wrong. They simply point out areas of the process that could be worthy of investigating further. The basic purpose of SPC is to reduce the variation. The two types of variation are ‘common cause’ and ‘special cause’ variation. Random variation is known as common cause variation while unexpected events/unplanned situations can result in special cause variation. Control charts are beneficial in identifying the types of variation. A process is considered ‘in control’ if it displays only common cause variation and ‘out of control’ if it shows special cause variation.

If sources of variation are detected and measured, they can then be corrected which may reduce the waste in production and improve the quality of the process. Assessing the causes of variation within a process
proves crucial when actions can be taken to improve process. In the 1950’s, Deming transformed post war Japan into the world leader of manufacturing with the effective application of SPC. Japanese products used to be regarded as 'cheap' and of 'poor quality'. This changed in July 1950 when Dr. Deming introduced statistical quality control to the Japanese industry, using the trademark colored-beads experiment [5]. This approach is gradually being applied in health care by considering healthcare systems as processes. In addition to providing a step for process improvement within health care, SPC charts are also alternative methods for displaying data visually over time.

**History of SPC and Previous Applications in Health Care**

Walter Shewhart, a doctor of physics, introduced the first control chart. He was employed by Western Electric, the manufacturing unit for American Telephone and Telegraph (AT&T), to “study the carbon microphone and to develop experimental techniques for measuring its properties” [6]. Shewhart, who had studied with Fisher in England, was greatly involved in statistical techniques and began to search ways to use them in the manufacturing of telephone parts. Shewhart noticed that there are limits to the random variation seen in most processes, even though there is variation in everything, and that these limits could be derived statistically. In 1924, he created the first process control chart: his procedure was to analyze and chart averages of small samples rather than individual values. Two of Shewhart’s colleagues at Western Electric, W. Edwards Deming and Joseph Juran, took his work further. In general, these statistical methods were not largely accepted until World War II, when the need for these techniques became prominent (Dr. Deming was very contributory in this effort).

SPC techniques are now being adopted not only in manufacturing processes but also in service-oriented domains such as health care, financial services, insurance and public services. Every year, Donald Berwick, M.D., president and CEO of IHI, states to “pledge allegiance to science and evidence” in IHI’s
annual conference. In other words, it’s time to use data and SPC to measure, monitor, and improve healthcare processes.

Thor et al reviews the literature of how statistical process control, with control charts used as the primary tool, has been applied in health care and presents the number of included articles by year of publication as shown in Figure 2. Among these reviewed articles, SPC has been applied in over 20 specialties or fields of healthcare such as anesthesia (intensive care), family practice (primary care), emergency medicine, cardiac surgery, and cardiology. Some of the variables used for SPC are blood pressure, heart rate, blood glucose measurements, daily pain scale recordings, patient fall rate, days between asthma attacks, time between patient check-in, average length of stay, proportion of low birth weight infants, number of medication errors and so on [7-9]. SPC methods are well applied in health care in the manner that nearly all healthcare systems can be viewed as processes that present inherent variability over time and provide a thorough introduction to the use and interpretation of SPC charts in a wide variety of healthcare applications [10, 11]. Use of control charts is often recommended to monitor and improve hospital performance [12].

Figure 2. The number of published articles on SPC in health care. Black bars: studies conducted in the USA; grey bars: studies outside the USA
SPC charts are chronological displays of process data which is used to understand the process statistically. The general format of a Shewhart-type control chart is given in Figure 3. The observed values of the process, for example the number of days between infections, are plotted on the chart over time. Three lines are also plotted, which are called the upper control limit (UCL), lower control limit (LCL), and center line (CL). These control limits are calculated statistically and used to define central tendency and to make a determination about stability of the process. Values that fall outside the limits (out of control) are strong evidence to support that there are non-systematic causes which should be investigated. Non-random behaviors between the limits also should not to be present, such as cycles, shifts from the center line, and trends. For detailed discussions, see Duncan [13], Grant [14], and Benneyan [10, 11, 15, 16].

![Figure 3. General format of a control chart](image)

To observe the changes in a process, different types of control charts have been developed such as Shewhart np and p charts for discrete outcomes from binomial processes, c and u charts for count data produced by Poisson processes, $\bar{X}$ and S charts for normally distributed continuous data [17]. To monitor the total or average number of events between adverse events, a g or h control chart [18-20], respectively, are developed by Kaminsky et al [19] and shown to outperform the p charts. The authors propose the number of procedures, events or days between infections as an alternate measure due to more timely
feedback, simplicity and almost instant availability of each observation. As a few recent examples of these measures, Nathanson et al [21] examined a 24-bed medical/surgical ICU experience, Walberg et al [22] analyzed the number of days between Pseudomonas aeruginosa infection, and Alemi et al [23] used time between control charts in monitoring asthma attacks. The sample unit of the process is also a factor for the choice of type of chart. For instance, if the process is set to control small variations and the sample has an individual unit, to use an EWMA or CUSUM control chart is advised [24].

In addition to these univariate control charts, the simultaneous control of two or more variables is usually referred in the literature as multivariate quality control. In 1947, Hotelling proposed a statistic which uniquely lends itself to plotting multivariate observations. This statistic, named Hotelling's $T^2$, is widely used in the industry as a multivariate process monitoring and control procedure. Another multivariate method is chi-square control chart which is also referred to as generalized $p$ chart in the literature [25]. After Shewhart suggested the use of chi-square test as a fifth criterion to detect the changes in a process in 1931, modeling with the chi-square control chart has been broadly studied. The basic idea of this chart is well laid out originally by Duncan (1950), Marcucci (1985), Nelson (1987), and Montgomery (2005) [4, 25-27].

**Brief Description of the Thesis**

This thesis consists of four chapters. The first chapter describes recent work to explore the best application of statistical process control methods to a state-wide neonatology quality improvement collaborative, called NeoQIC, including the comparative performance of statistical methods to detect changes in key physiologic data. The second chapter observes the chart performances via their average run lengths by constructing Shewhart charts and EWMA $g$ chart for several randomly generated series as well as series with small shifts in the process rate. While the first chapter is an application of SPC
methods to monitor adverse events, the second chapter investigates the ways to properly monitor adverse events.

The third chapter describes the Specialty Care project carried out with the Veterans Health Administration (VHA). This project mainly includes building a quality dashboard, the Specialty Care Compass, intended for use by the VHA as a part of the Specialty Care (SC) Collaborative. It includes practical and basic measures of access, care coordination and efficiency in an accessible manner for clinical teams. This tool targets several stakeholder groups: specialist providers and teams, specialty chiefs, partner providers, and local and national leadership.

The main purpose of the fourth and last chapter is to extend the application of control charts to cases where the data set is given in frequency format and there is no information about actual observation values, sample means and sample standard deviations. The question that is raised in this particular situation is how a sequence of these frequencies is properly monitored with the lack of actual observations and sample information. The Specialty Care dashboard tool described in the previous chapter also includes this type of data sets. A case study is also presented as an illustration for this notion with a data set from colonoscopy services of one of the VA facilities. The methods discussed in this chapter can be powerful tools for monitoring the grouped data sets which would lead to more efficient results for the Collaborative.
Chapter 1. Applications of SPC to Neonatology Quality Improvement

1. Background

In this project, we describe recent work to explore how to best apply statistical process control methods to a state-wide neonatology quality improvement collaborative, called NeoQIC, including the comparative performance of statistical methods to detect changes in key physiologic data.

SPC methods are a good fit to apply in health care in the manner that nearly all healthcare systems can be viewed as processes that present inherent variability over time and provide a thorough introduction to the use and interpretation of SPC charts in a wide variety of healthcare applications [10, 11]. Use of control charts is often recommended to monitor and improve hospital performance [12]. SPC methods have been used in healthcare improvement studies in various fields such as anaesthesia, family practice, emergency medicine, cardiology, nursing, mental health, clinical chemistry, and so on [7-9]. An important application area of SPC in health care is neonatology which focuses primarily on the medical needs of newborns. The purpose of this paper is to illustrate the use of appropriate control charts and the impact of misuse on the power to detect changes for healthcare data provided by the Neonatal Quality Improvement Collaborative of Massachusetts. To conduct the analysis, software packages Excel and Minitab are used.

NeoQIC is a voluntary organization of healthcare providers committed to improve the health of newborns. Its mission is to promote a continuous quality improvement culture [28]. The data includes body temperature measurements and infection information obtained from 500 infants born between the years 2006 and 2009 over 48 months, i.e. there are 48 sub-groups with varying sample sizes that sum up to 500. Infection variable, however, has 19 missing values which are ignored during the analysis of the data. There are four variables, month, year, temperature, and infection, each of which is explained in Table 1.
Table 1. Variables and their definitions

<table>
<thead>
<tr>
<th>Variables</th>
<th>Definitions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Month</td>
<td>Birth month of an infant</td>
</tr>
<tr>
<td>Year</td>
<td>Birth year of an infant between 2006 and 2009</td>
</tr>
<tr>
<td>Temperature</td>
<td>Body temperature of an infant</td>
</tr>
<tr>
<td>Infection</td>
<td>Binary variable indicating whether an infant has an infection or not. No infection corresponds to 0 while infection existence corresponds to 1.</td>
</tr>
</tbody>
</table>

2. Methodology

Statistical investigation of data usually starts with explanatory data analysis which enables to understand the general behavior of the variables using tools such as scatter plots, probability plots, histograms, autocorrelation functions, and normality tests. While the statistics and principles underlying SPC are relatively straightforward, determining the correct chart to use is especially important in terms of achieving efficient results. The choice of the control chart is first based on the type of data (whether it is variable or attributes) and next issues such as purpose of constructing the chart, subgroup sizes, and so on. In this analysis, temperature variable is continuous and infection variable is binary. Because the process parameters are estimated from the data, the standards not given charts are developed.

To monitor the number of infections observed in a month, $p$ chart and standardized $p$ charts are constructed. Due to the unequal subgroup sample sizes, UCL and LCL values of a $p$ chart are not constant and calculated for each subgroup. Another way to deal with unequal sample sizes is to construct a standardized $p$ chart, - with center line 0, upper and lower control limits +3 and -3, respectively - for the variable:
\[ Z_i = \frac{\hat{p}_i - \hat{p}}{\sqrt{\frac{\hat{p}(1 - \hat{p})}{n_i}}} \]  

(1)

For continuous temperature variable \( T \), the individuals (XMR), exponentially weighted moving average (EWMA), and cumulative sum (CUSUM) control charts are obtained in addition to \( \bar{X} \) and \( R \) and \( \bar{X} \) and \( S \) control charts. While constructing the XMR chart, a moving range of \( n = 2 \) observations is used and is calculated as

\[ MR_i = |x_i - x_{i-1}|, \]  

(2)

where \( i = 1 \ldots n \) [4]. Both EWMA and CUSUM charts include information from all sequence of observations and have more power to detect small shifts. EWMA statistic can be calculated as

\[ EWMA_i = \lambda x_i + (1 - \lambda)(EWMA)_{i-1}, \]  

(3)

where \( EWMA_0 \) is the sample average of preliminary data, \( \mu_0 = \bar{x} \) [29]. The 3\( \sigma \) control limits are calculated using a weighting factor of \( \lambda = 0.2 \). The one-sided upper and lower CUSUM statistics, \( C^+ \) and \( C^- \), respectively monitor positive and negative shifts in the process and are calculated as

\[ C^+_i = \max[0, x_i - (\mu_0 + K) + C^+_{i-1}] \]  

and \( C^-_i = \max [0, (\mu_0 - K) - x_i + C^-_{i-1}] \),

(4)

where the starting values \( C^+_0 \) and \( C^-_0 \) are 0 [30], \( x_i \) is the \( i^{th} \) observation on the process, \( \mu_0 \) is the target value taken as the sample average of the temperature data, and \( K \) is the reference value obtained as \( \frac{3\sigma_0}{2} \) and \( \sigma \) is the sample standard deviation [30]. Furthermore, \( g \) chart is constructed to monitor the number of healthy infants between infections, rather than the number of events per time period.

A common approach of monitoring continuous measures in health care is to use a \( p \) chart for a binary variable created by comparing the continuous outcome measures to a threshold value chosen based on expertise in the field. For an infant, for example, having the body temperature lower than 36°C or 36.5°C is considered risky and \( p \) charts are used to monitor the proportion of infants with body temperatures
lower than these thresholds. To investigate the impact of this approach on detecting the changes in the infant body temperature, the continuous temperature variable is converted to a binary variable $T^*$ as

$$T^* = \begin{cases} 
1 & \text{if } \text{Temperature} < 36^\circ\text{C} \\ 
0 & \text{if } \text{Temperature} \geq 36^\circ\text{C} 
\end{cases}$$  \hspace{1cm} (5)

This study also compares the power of monitoring this new binary variable via a $p$ chart versus monitoring the original continuous measure, temperature, via an $\bar{X}$ chart. The performance of these $\bar{X}$ and $p$ charts are compared using operating characteristic (OC) and average run length (ARL) curves, where an OC curve illustrates the probability of not detecting a shift, i.e.

$$\beta = P(LCL \leq \theta \leq UCL|\theta_1 = \theta + \delta \theta),$$  \hspace{1cm} (6)

where $\theta$ is the population parameter. ARL is the expected number of samples taken before an out of control signal is obtained, and is computed as

$$\text{ARL} = \frac{1}{1-\beta}.$$  \hspace{1cm} (7)

The relationship between the temperature of infants and the probability of developing an infection furthermore is investigated by developing a binary logistic regression model.

3. Results

3.1. $p$ chart for infection rate

Standards not given $p$ charts are constructed for the proportion of infants that develop infection. Figure 4(a) and (b) respectively illustrate the $p$ and standardized $p$ charts both of which indicate statistical out of control state at 46th month for the proportion of infants that develop infection.
Figure 4. (a) p chart and (b) standardized p chart for infection data

3.2. \( g \) chart for the number between infections

The \( g \) chart developed by Benneyan [18-20] is a powerful tool for surveillance of the number of events between infections or other rare events that follow geometric or negative binomial distributions. The \( g \) chart in Figure 5 is constructed by using number-until type formulas with minimum variance unbiased estimator (MVUE) [19] and indicates no unusual behavior in the number of healthy infants until the 1st infection is observed. The chart also shows that UCL is wider when maximum likelihood estimator (MLE) is used instead of MVUE to estimate the proportion.

3.3. \( \bar{X} \) and \( S \) control charts for temperature

\( \bar{X} \) and \( R \) and \( \bar{X} \) and \( S \) charts are constructed with normality and independence assumptions for the mean body temperature, where \( \bar{X} \) is the average of each subgroup, \( R \) the range calculated by subtracting the
minimum observation from the maximum in each sample, and \( S \) the sample standard deviation. Figure 6 presents \( \bar{X} \) and \( S \) charts depicting the change of the variability after 35\(^{th}\) period. The \( \bar{X} \) and \( R \) charts also display similar results as expected.

![Graphs showing \( \bar{X} \) and \( S \) charts for temperature](image)

Figure 6. (a) \( S \), (b) \( \bar{X} \) with \( S \) charts for the temperature variable

### 3.4. XMR, EWMA and CUSUM charts for temperature

The moving range and individual charts in Figure 7(a) and (c), respectively, agree with the above Shewhart charts. However, both charts have many more out of control signals than the \( R \) (\( S \)) and \( \bar{X} \), respectively. This might be due to the reason that these charts are not robust to a possible non-normality. EWMA and CUSUM charts, which are more robust to non-normality than any of the above control charts, are also constructed for further investigation of the temperature data. Both charts, presented in Figure 7(b) and (d), respectively, detect the drop in the average body temperature of the infants as early as the 35\(^{th}\) month whereas \( \bar{X} \) chart fails to detect until the 46\(^{th}\).
Figure 7. (a) Individual chart, (b) EWMA, (c) MR, and (d) CUSUM control charts for the temperature variable

3.5. The relationship between temperature and infection variables

To assess whether there is a significant relationship between body temperature of the infants and the infection data, simple binary logistic regression is applied. The model is given by

\[ \text{logit}(p_i) = \ln \left( \frac{p_i}{1 - p_i} \right) = \beta_0 + \beta_1 x_{1i}, \]  

(8)

where \( p_i \) is the fraction of infection, the response variable, and the independent variable \( x_1 \) is the body temperature of the infant. The Minitab output in Table 2 indicates that the coefficients \( \beta_0 \) and \( \beta_1 \) are not significant at the 0.05 level. One then can conclude that there is not enough statistical evidence to indicate a significant relationship between infection and temperature data.
Table 2. Binary logistic regression for the infection and temperature variables

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coef</th>
<th>SE Coef</th>
<th>Z</th>
<th>P</th>
<th>Ratio</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>10.0464</td>
<td>7.27097</td>
<td>1.38</td>
<td>0.167</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TEMPERATURE</td>
<td>-0.347783</td>
<td>0.202707</td>
<td>-1.72</td>
<td>0.086</td>
<td>0.71</td>
<td>0.47</td>
<td>1.05</td>
</tr>
</tbody>
</table>

3.6. $p$ chart for $T^*$

This chart is operated by plotting the sample fraction $\hat{p}_i$, $i = 1, \ldots, 48$, of infants whose body temperatures are under the threshold 36 as given in Equation 5, for each monthly sample. Figure 8(a) and (b) illustrate the $p$ and standardized $p$ charts. Although there is no out of control signal in any chart, an upward shift is observed in the proportions of infant temperatures after the 35th point corresponding the 11th month of 2008. To further investigate whether there is a difference in the mean temperatures of infants before ($\mu_x$) and after ($\mu_y$) this time period, the data is divided into two samples of sizes 354 and 146. A one sided $t$-test for unequal variances (according to an $F$ test with p-value 0.005) then is performed to determine if the sample mean of $X$ is greater than of $Y$. The test indicates enough evidence (p-value: 0.00049) to conclude that the sample mean of $X$ is less than the sample mean of $Y$.

![Figure 8](image-url)
3.7. Relative performances of the charts

The OC and ARL curves in Figure 9(a) and (b) assess the ability of $\bar{X}$ and $p$ charts to detect a shift in the mean body temperature of infants such that $\mu_1 = \mu_0 + \delta \sigma$. Both curves indicate that the $\bar{X}$ chart is more effective than the $p$ chart in detecting a small shift – approximately less than 1.5$\sigma$ – on the first sample following the shift. According to the OC curve, the probability of not detecting a change of magnitude 0.7$\sigma$, for example, is $\beta = 0.78$. Thus, the probability that the shift will be detected is $1-\beta = 0.22$. The expected number of samples to detect a shift of 0.7$\sigma$ in the average body temperature of an infant, ARL is $1/(1-0.78) = 4.6$, approximately 5. On the other hand, the probability that $p$ chart constructed for the binary variable $T^*$ detects a shift of the same magnitude is approximately $1-\beta = 1-0.93 = 0.07$ with a corresponding ARL of 15, implying that using the $p$ chart for a continuous measure causes delays in detection of a shift in the mean.

![OC curves](image1)

![ARL curves](image2)

Figure 9. (a) OC curves and (b) ARL curves for $\bar{X}$ chart and $p$ chart of temperature

4. Discussion

This studyillustrates the use and performance of SPC charts in monitoring health outcomes of infants in a neonatal process taken from NeoQIC. Both Shewhart and advanced individuals control charts are used for better understanding of the variability in the body temperature data. There is enough reason to believe that the variation in the body temperatures of the infants has increased and the average body temperature has
dropped especially in the last year of the 2006-2009 period. The results indicate that while the individuals charts XMR, EWMA, and CUSUM can detect anomalies in the data faster than an $\bar{X}$ chart, XMR chart is less robust to non-normality than EWMA and CUSUM, leading to more false alarms. The $p$ chart constructed for the binary variable $T^*$ performs worse than the $\bar{X}$ chart for $T$ and fails to detect the drop in the average body temperature of the infants. The consequences of using an attributes chart instead of a variables chart for a continuous measure is highlighted via OC and ARL curves developed for both charts indicating power loss and delays in detection of changes in the system. As future work, the performances of EWMA versus CUSUM charts can be observed.
Chapter 2. Relative Performance of Shewhart and EWMA $p$ and $g$ Statistical Control Charts for Monitoring Adverse Healthcare Event Rates

1. Background

Shewhart $p$ and $g$ statistical control charts often are used to detect special cause variation in healthcare adverse event (AE) rates and other performance measures, the latter being used for the number of cases between $n$ adverse events (where usually $n = 1$). Example applications include surgical site infections, ventilator associated pneumonia, care compliance, and others. For low rates $g$ charts based on underlying negative binomial distributions outperform conventional binomial based $p$ charts. When change detection run lengths for either chart are still too slow, optimizing sample sizes or the use of exponentially weighted moving average (EWMA) version can further improve performance. We investigate the relative performance (average run lengths) of these charts across a range of different scenarios, baseline AE rates, and chart design parameters, including the result that optimal change detection performance for $g$ charts often is maximized for $n \geq 2$ (that is, cases between every second AE).

To monitor the total or average number of events between adverse events, the $g$ and $h$ control charts respectively, are developed by Kaminsky et al [19] and shown to outperform the $p$ charts. The authors propose the number of procedures, events or days between infections as an alternate measure due to more timely feedback, simplicity and almost instant availability of each observation. As a few recent examples of these measures, Nathanson et al [21] examined a 24-bed medical/surgical ICU experience, Walberg et al. [22] analyzed the number of days between Pseudomonas aeruginosa infection, and Alemi et al [23] used time between control charts in monitoring asthma attacks.

To monitor a process rate over time, a $p$ chart based on binomial distribution can be developed. For a sample coming from geometric distribution and having a low proportion, it is desirable to construct a $g$
chart to monitor the adverse events. The sample unit of the process is also a factor for the choice of type of the chart. For instance, if the process is set to control small variations and the sample has an individual unit, to use EWMA or CUSUM control chart is advised [24]. In this situation, it is also possible to use an EWMA chart to detect the small variations.

In this study, it is intended to observe the chart performances by constructing Shewhart charts and EWMA $g$ chart for several in-control series as well as series with small shifts in the process rate. The series are obtained from geometric distribution for given process rates. The adverse event proportion is altered and the performance of each control chart is observed. The criterion used to evaluate the performance of the charts is Average Run Length (ARL) indicating how many samples are expected to be taken in order to detect an out-of-control signal [10]. For an in-control process, ARL should be large corresponding to a low false alarm rate. In contrast, for an out-of-control process, it would be small implying that the change is detected quickly. The out-of-control ARLs of the two charts are used to compare the chart performances. For ARL calculations of the $g$ charts, $k$-sigma control limits and probability-based limits are used and for EWMA chart ARLs are obtained via Monte Carlo simulation. When the control charts perform slowly to detect the changes, optimal samples sizes and parameter designs are advised.

2. Methodology

2.1. $p$, $g$, and EWMA $g$ charts

The $p$ chart is a popular choice to plot the proportion of defective items in a process per subgroup. The data collected are the number of defective items in each subgroup, which is assumed to follow a binomial distribution with a proportion parameter ($p$). The control limits for a given $p$ and sample size $n_i$ can be easily calculated with the equations in (9). The subgroup size for $p$ chart is determined by one general rule of thumb [15] where $n$ should be greater than or equal to $[3/p]^+$. 
\[
Center \ Line \ (CL) = p
\]

\[
Upper \ Control \ Limit \ (UCL) = p + k \frac{p(1-p)}{n_i}
\]

\[
Lower \ Control \ Limit \ (LCL) = p - k \frac{p(1-p)}{n_i}
\] (9)

The Shewhart-type \( g \) and \( h \) control charts were developed and investigated by Kaminsky and Benneyan [19] which have shown to exhibit better statistical operating characteristics over traditional \( p \) charts, especially when the rate of occurrence is very low [10, 18, 19]. These charts are based on geometric and negative binomial distributions as alternate control charts. When dealing with geometric random variables and low frequency data, this type of charts are efficient tools to obtain more accurate results.

As a brief motivation, \( g \) chart takes into account the number of Bernoulli trials until or before the first failure which is a geometric distribution with the probability distribution function

\[
P(X = x) = p(1-p)^{x-a}, \quad x = a, a+1, a+2, ...
\] (10)

where \( a = 1 \) for number-\until and \( a = 0 \) for number-\before data. The sum of \( n \) independent geometric random variables, \( T = X_1 + X_2 + \cdots + X_n \), has negative binomial distribution with probability function

\[
P(T = t) = \binom{n(1-a)+t-1}{n-1} p^n(1-p)^{t-na}, \quad t = na, na+1, na+2, ...
\] (11)

The expected value and variance of the negative binomial distribution lead to the control limits in Table 3 [19, 31]. Note that these formulas are for standards given case which is used in this study. Rate estimated cases with maximum likelihood estimator and minimum variance unbiased estimator are also can be found in Kaminsky et al [19]. Through the study, the data is assumed to be number-\until type \( (a = 1) \) for all control charts.
The relative performance comparisons in the study are based on the ARLs and percent in-control. The lower control limit (LCL) is taken as $na$ because of the nonnegative nature of geometric distribution which resulted in little-to-no power to detect increases in the rate. Several design approaches discussed by Benneyan [20] for this case are applied to optimize the performance of Shewhart $g$ charts. Narrower limits were obtained with $k = z_{\alpha=0.05} = 1.96$ standard deviation limits as well as probability-based control limits with $\alpha = 0.05$ false alarm rate. Table 3 contains the probability-based control limits of $g$ chart when the sample size $n$ is 1 and the cases for $n > 1$ are then calculated iteratively.

The next step involves the EWMA chart which includes information from all sequence of observations and therefore has more power to detect small shifts in the process rate than Shewhart charts. EWMA control chart was introduced by Roberts [32] in 1959 as an alternative to Shewhart charts. Hunter [33] pointed out that EWMA assigns more weight to recent information while Shewhart charts emphasize the last defined point. The properties of EWMA charts were investigated by Crowder [34] and Lucas and Saccucci [35] analytically and its performance was compared in several studies [36-39].

The ARLs of EWMA chart are obtained by generating random numbers from geometric distribution with 1000 iterations because 1000 iterations are enough to get the same values for EWMA with $\lambda = 1$ and $k$-sigma limit $g$ chart. The EWMA statistic is calculated as

Table 3. Control limits for $g$ charts

<table>
<thead>
<tr>
<th></th>
<th>$k$ sigma-type limits</th>
<th>Probability-based limits $(n = 1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Centerline</strong></td>
<td>$CL = n\left(\frac{1-p}{p} + a\right)$</td>
<td>$CL' = \left[\ln(0.5) \over \ln(1-p) + a\right]^{+}$</td>
</tr>
<tr>
<td><strong>Upper Control Limit</strong></td>
<td>$UCL = n\left(\frac{1-p}{p} + a\right) + k \sqrt{\frac{n(1-p)}{p^2}}$</td>
<td>$UCL' = \left[\ln(\alpha_{UCL}) \over \ln(1-p) + a - 1\right]^{+}$</td>
</tr>
<tr>
<td><strong>Lower Control Limit</strong></td>
<td>$LCL = n\left(\frac{1-p}{p} + a\right) - k \sqrt{\frac{n(1-p)}{p^2}}$</td>
<td>$LCL' = \left[\ln(1-\alpha_{LCL}) \over \ln(1-p) + a\right]^{-}$</td>
</tr>
</tbody>
</table>
\[ \text{EWMA}_i = \lambda x_i + (1 - \lambda)(\text{EWMA})_{i-1} \]  

(12)

where \( \text{EWMA}_0 = \mu_0 \) is the mean of geometric distribution \([35]\). The choice for the parameter \( \lambda \) is explained in the next section.

### 2.2. Choosing the parameters of EWMA chart

Montgomery \([4]\) states that values of \( \lambda \) in the interval \( 0.05 \leq \lambda \leq 0.25 \) work well in practice and a good rule is to use smaller values of \( \lambda \) to detect smaller shifts. Table 4 includes the ARLs of EWMA chart for the popular choices \( \lambda = 0.05, 0.1, \) and \( 0.2 \) when \( k = 3 \) or \( k = z_{\alpha=0.05} = 1.96, \) and \( p = 0.01. \)

<table>
<thead>
<tr>
<th>Table 4. Control limits for EWMA chart</th>
</tr>
</thead>
<tbody>
<tr>
<td>CL</td>
</tr>
<tr>
<td>UCL</td>
</tr>
<tr>
<td>LCL</td>
</tr>
</tbody>
</table>

For this study, \( \lambda \) is chosen as 0.05 since we are dealing with small shifts. In addition, we would like to have a large ARL when the process is in-control and small ARL when it is out-of-control. The EWMA chart with \( \lambda = 0.05 \) has the largest in-control ARL and detects the positive shifts in the rate faster. In many applications, the most popular choice for \( \lambda \) is 0.2 which gives quite high ARLs for positive shifts. The control limits used for the chart are given in Table 4 where \( \sigma \) is the standard deviation of geometric distribution. The power of the chart is then observed for shifted process rates and compared to the Shewhart charts.
Table 5. ARLs of EWMA chart for different $\lambda$ and $k$ values for $p = 0.01$

<table>
<thead>
<tr>
<th>Shift</th>
<th>$\lambda = 0.05$</th>
<th>$\lambda = 0.1$</th>
<th>$\lambda = 0.2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$k = 3$</td>
<td>$k = 1.96$</td>
<td>$k = 3$</td>
</tr>
<tr>
<td>0</td>
<td>866.913</td>
<td>107.761</td>
<td>410.436</td>
</tr>
<tr>
<td>0.1</td>
<td>3135.72</td>
<td>110.754</td>
<td>1335.85</td>
</tr>
<tr>
<td>0.2</td>
<td>1782.32</td>
<td>73.662</td>
<td>4168.45</td>
</tr>
<tr>
<td>0.3</td>
<td>637.277</td>
<td>54.57</td>
<td>9077.68</td>
</tr>
<tr>
<td>0.4</td>
<td>278.42</td>
<td>39.278</td>
<td>5980.62</td>
</tr>
<tr>
<td>0.5</td>
<td>161.518</td>
<td>30.392</td>
<td>2674.85</td>
</tr>
<tr>
<td>0.6</td>
<td>105.268</td>
<td>24.755</td>
<td>1238.89</td>
</tr>
<tr>
<td>0.7</td>
<td>75.616</td>
<td>20.935</td>
<td>649.424</td>
</tr>
<tr>
<td>0.8</td>
<td>59.601</td>
<td>18.785</td>
<td>384.679</td>
</tr>
<tr>
<td>0.9</td>
<td>49.716</td>
<td>16.444</td>
<td>243.044</td>
</tr>
</tbody>
</table>

2.3. Performance Analysis

After determining the limits for a given process rate ($p$), the probability that a point is inside the limits is calculated as in equation (13) for $p$ chart which is based on the binomial distribution.

\[
P(\text{no signal}) = P(LCL \leq X \leq UCL) = \sum_{x=[LCL]}^{[UCL]} B(x; n, p)
\]

\[
= \sum_{x=[LCL]}^{[UCL]} \binom{n}{x} p^x (1 - p)^{n-x}
\]

The same probability for Shewhart $g$ chart is obtained as in equation (14) which is based on the cumulative distribution function of negative binomial distribution. The failure number, $n$, is initially set to 1 as the first failure and effect of increasing $n$ is also discussed.
\[ P(\text{no signal}) = P(\text{LCL} \leq X \leq \text{UCL}) = \sum_{x=[\text{LCL}]}^{[\text{UCL}]} \text{NB}(x; n, a, p) \]

\[ = \sum_{x=[\text{LCL}]}^{[\text{UCL}]} x \left( \frac{n(1-a) + x - 1}{n-1} \right) p^n (1-p)^{x-na} \]  \quad (14)

Similarly, for \( g \) chart with probability-based limits, the probability of the random variable \( X \) being less than lower control limit and greater than upper control limit is

\[ \beta_{\text{LCL}} = P(X < \text{LCL}') = 1 - (1-p_a)^{[\text{LCL}']^+ - a} \quad \text{and} \]

\[ \beta_{\text{UCL}} = P(X > \text{UCL}') = (1-p_a)^{[\text{UCL}']^- - a+1} \]  \quad (15)

where \([\ ]^-\) indicates the round-down function, \([\ ]^+\) is the round-up function and \( p_a \) is the out-of-control rate intended to detect. More information on the derivation can be found in [20]. Note that the limits of the probability-based limit \( g \) chart are only for \( n = 1 \) and the desired false alarm rates are set to \( \alpha_{\text{LCL}} = \alpha_{\text{UCL}} = \alpha/2 = 0.025 \). ARLs are related to probability of no signal as

\[ \text{ARL} = \frac{1}{1 - P(\text{no signal})} \]  \quad (16)

After obtaining in-control ARLs for a given process rate, \( p \), the performances of the control charts are observed by applying small positive shifts (\( \delta \)) to the rate which can be defined as \( p_a = p + p\delta \). The results are displayed with ARL and probability in-control curves in the next section.
3. Results

3.1. Shewhart $p$ vs $g$ chart for low rates

The $g$ chart has proven to be an efficient tool to monitor the number of cases between hospital-acquired infections or other adverse events. On the other hand, EWMA chart is known to be well performing for detecting the small shifts in a process. The main focus of this study is maximizing the performance of Shewhart $g$ chart in the case of small shifts and very low adverse event rate as well as comparing with $p$ chart and EWMA chart versions.

![Figure 10](image_url)

(a) Probability in-control of $k = 3$ sigma $p$ and $g$ charts for positive rate changes.

(b) Comparison of $g$ and $p$ charts by average number of items until a signal

It has been shown that the $g$ chart outperform the $p$ chart especially for low rates [10, 18, 19]. This is also illustrated in Figure 10(a) for two different infection processes having in-control probabilities of $p = 0.1$ and $p = 0.01$. For $p = 0.1$, the optimal sample size for $p$ chart is 30 and for $p = 0.01$, sample size of 300 is required. The ARLs for $g$ chart are obtained with the same sample sizes. Although the $p$ chart is performing well for $p = 0.1$, the performance of $g$ chart becomes better as the infection rate decreases. Even for $p = 0.00001$, $p$ chart would require a large amount of sample size unlike $g$ chart.
A much more suitable metric for comparing the $p$ and $g$ control charts is the average number of items (ANI) until an out-of-control signal [20]. This metric considers the number of items inspected, whereas the ARL metric only takes into account the number of times a chart statistic is plotted. Determining the best methods of performance is based on having the smallest out-of-control ANI values across a range of shifts in $p$. Figure 10(b) compares the ANI values for the $p$ chart ($n = 300$ with the rule of thumb $[3/p]^+$ when $p = 0.01$) and $g$ chart ($n = 1$) for the rate of 0.1 and 0.01. The figure illustrates that the $g$ chart is ineffective for detecting rate increases, while the $p$ chart is ineffective for detecting rate decreases for both rates. Both charts perform better when the rate is higher ($p = 0.1$).

### 3.2. Improving $g$ charts

#### 3.2.1. EWMA vs Shewhart $g$ charts and probability-based limits

In his study, Benneyan compares the conventional $k = 3$ standard deviation control limits to $k = z_{\alpha=0.05} = 1.96$ standard deviation warning limits and $\alpha = 0.05$ probability limits. It is also illustrated in Table 6 that when the in-control rate is 0.1 or 0.00001, the $g$ chart with $k = 1.96$ detects the positive shifts in the rate better than the $g$ chart with $k = 3$. For the low rate 0.0001, the probability-based limits tend to result in the best ARLs. The probability limits are also better than EWMA chart to detect the small shifts for the low rate. However, the in-control ARL of EWMA is much greater than the others. Therefore we can say that to use EWMA chart is more powerful to detect the shifts that are not very small.

The curves in Figure 11 representing probability of being inside the control limits in $y$-axis and the shift in the infection rate in $x$-axis also suggest the same conclusion. Note the improvement of $g$ chart with and $\alpha = 0.05$ probability limits over the $k = 1.96$ sigma limits having actually the same false alarm rate (0.05) but failing to detect rate increases.
Table 6. Average Run Lengths for EWMA and $g$ charts with different scenarios for $n = 1$

<table>
<thead>
<tr>
<th>Shift</th>
<th>EWMA chart</th>
<th>Shewhart $g$ chart</th>
<th>EWMA chart</th>
<th>Shewhart $g$ chart</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$k = 3$</td>
<td>$k = 1.96$</td>
<td>$k = 3$</td>
<td>$k = 1.96$</td>
</tr>
<tr>
<td>0</td>
<td>897.97</td>
<td>107.95</td>
<td>54.8</td>
<td>19.11</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>44.39</td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>3338.93</td>
<td>113.63</td>
<td>83.79</td>
<td>26.13</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>66.37</td>
<td></td>
</tr>
<tr>
<td>0.2</td>
<td>1501.33</td>
<td>73.37</td>
<td>128.72</td>
<td>35.85</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>99.68</td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td>535.1</td>
<td>49.61</td>
<td>198.73</td>
<td>49.37</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>150.42</td>
<td></td>
</tr>
<tr>
<td>0.4</td>
<td>230.16</td>
<td>34.49</td>
<td>308.36</td>
<td>68.24</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>228.06</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>133.75</td>
<td>27.38</td>
<td>480.93</td>
<td>94.68</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>347.47</td>
<td></td>
</tr>
<tr>
<td>0.6</td>
<td>88.38</td>
<td>22.85</td>
<td>754.03</td>
<td>131.88</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>532.04</td>
<td></td>
</tr>
<tr>
<td>0.7</td>
<td>67.82</td>
<td>20.09</td>
<td>1188.59</td>
<td>184.42</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>818.82</td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td>54.2</td>
<td>17.64</td>
<td>1883.97</td>
<td>258.95</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1266.78</td>
<td></td>
</tr>
<tr>
<td>0.9</td>
<td>44.81</td>
<td>15.52</td>
<td>3003.1</td>
<td>365.11</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1970.33</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Shift</th>
<th>EWMA chart</th>
<th>Shewhart $g$ chart</th>
<th>EWMA chart</th>
<th>Shewhart $g$ chart</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$k = 3$</td>
<td>$k = 1.96$</td>
<td>$k = 3$</td>
<td>$k = 1.96$</td>
</tr>
<tr>
<td>0</td>
<td>871.7</td>
<td>113.12</td>
<td>54.6</td>
<td>19.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>3194.55</td>
<td>119.26</td>
<td>81.45</td>
<td>25.95</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>22.35</td>
<td></td>
</tr>
<tr>
<td>0.2</td>
<td>1798.25</td>
<td>77.88</td>
<td>121.51</td>
<td>34.88</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>23.88</td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td>627.48</td>
<td>51.04</td>
<td>181.27</td>
<td>46.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>24.61</td>
<td></td>
</tr>
<tr>
<td>0.4</td>
<td>281.11</td>
<td>39.45</td>
<td>270.43</td>
<td>63.06</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>24.67</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>164.86</td>
<td>30.1</td>
<td>403.43</td>
<td>84.78</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>24.27</td>
<td></td>
</tr>
<tr>
<td>0.6</td>
<td>103.72</td>
<td>25.39</td>
<td>601.86</td>
<td>113.98</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>23.57</td>
<td></td>
</tr>
<tr>
<td>0.7</td>
<td>80.97</td>
<td>21.55</td>
<td>897.87</td>
<td>153.24</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>22.72</td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td>62.79</td>
<td>18.6</td>
<td>1339.47</td>
<td>206.03</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>21.81</td>
<td></td>
</tr>
<tr>
<td>0.9</td>
<td>50.91</td>
<td>16.87</td>
<td>1998.26</td>
<td>277</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>20.9</td>
<td></td>
</tr>
</tbody>
</table>

Figure 11. Probabilities of being in-control for $n = 1$ when (a) $p = 0.1$ and (b) $p = 0.00001$
3.2.2. **Optimal sample size**

Another way to improve the performance of the charts is to take large enough subgroups so that LCL also has a detection power. To have an LCL greater than the minimum possible subgroup total, \( na \), the subgroup size \( n \) should be

\[
n > \frac{k^2}{1 - p}
\]

for \( n \geq 1 \) [20]. Benneyan shows that the probability limits still result in better sensitivity even when the required sample size is set and \( k \)-sigma LCL > \( na \). This can be also seen in Table 7 where the sample size is set to 10 for both \( p = 0.1 \) and \( 0.00001 \). The probability in-control curves in Figure 12 also indicate that although \( n \) is adjusted to improve the performance of \( k \)-sigma \( g \) chart, the probability limits and EWMA chart still perform better. Note that, also, the EWMA chart has much higher in-control ARL and is able to detect the positive shifts as quick as probability-based limit \( g \) chart. This leads us to conclude that using EWMA chart will result in more power to detect the rate increases.

<table>
<thead>
<tr>
<th>Shift</th>
<th>( p = 0.1 )</th>
<th>( p = 0.00001 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( g ) chart ( (k = 3) )</td>
<td>( g ) chart ( (\alpha = 0.05) )</td>
</tr>
<tr>
<td>0</td>
<td>1278.56</td>
<td>153.27</td>
</tr>
<tr>
<td>0.1</td>
<td>104.22</td>
<td>522.87</td>
</tr>
<tr>
<td>0.2</td>
<td>27.88</td>
<td>1942.89</td>
</tr>
<tr>
<td>0.3</td>
<td>15.39</td>
<td>7798.85</td>
</tr>
<tr>
<td>0.4</td>
<td>10.59</td>
<td>33594.22</td>
</tr>
<tr>
<td>0.5</td>
<td>7.89</td>
<td>154470.41</td>
</tr>
<tr>
<td>0.6</td>
<td>6.32</td>
<td>754947.58</td>
</tr>
<tr>
<td>0.7</td>
<td>5.31</td>
<td>3908221.78</td>
</tr>
<tr>
<td>0.8</td>
<td>4.73</td>
<td>21370710.48</td>
</tr>
<tr>
<td>0.9</td>
<td>4.20</td>
<td>123155945.7</td>
</tr>
</tbody>
</table>
In the literature, it is very common to set the parameters of the charts to get approximately the same in-control ARLs and apply shifts in order to observe the performances of the charts with the out-of-control ARLs. As Table 8 illustrates, the in-control ARL of the $g$ chart with $k = 3$ sigma limit is 54.80. For the EWMA chart, $k = 1.68$ sigma limits gives close in control ARL. When the charts are run with these limits, the EWMA chart is considerably more powerful for detecting the positive shifts in the rate. Setting $k$-value of $g$ chart according to in-control ARL of EWMA chart (897.97) results in the same conclusion.

Table 8. ARLs when in-control ARLs are common (0 shift represents in-control)

<table>
<thead>
<tr>
<th>Shift</th>
<th>$g$ chart ($k = 3$)</th>
<th>EWMA ($k = 1.68$)</th>
<th>$g$ chart ($k = 5.78$)</th>
<th>EWMA ($k = 3$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>54.80</td>
<td>55.22</td>
<td>848.16</td>
<td>897.97</td>
</tr>
<tr>
<td>0.1</td>
<td>83.79</td>
<td>56.13</td>
<td>1733.96</td>
<td>3338.93</td>
</tr>
<tr>
<td>0.2</td>
<td>128.72</td>
<td>43.10</td>
<td>3573.63</td>
<td>1501.33</td>
</tr>
<tr>
<td>0.3</td>
<td>198.73</td>
<td>31.12</td>
<td>7426.22</td>
<td>535.10</td>
</tr>
<tr>
<td>0.4</td>
<td>308.36</td>
<td>22.83</td>
<td>15563.21</td>
<td>230.16</td>
</tr>
<tr>
<td>0.5</td>
<td>480.93</td>
<td>18.81</td>
<td>32899.46</td>
<td>133.75</td>
</tr>
<tr>
<td>0.6</td>
<td>754.03</td>
<td>15.85</td>
<td>70165.83</td>
<td>88.38</td>
</tr>
<tr>
<td>0.7</td>
<td>1188.59</td>
<td>13.91</td>
<td>151008.74</td>
<td>67.82</td>
</tr>
<tr>
<td>0.8</td>
<td>1883.97</td>
<td>12.34</td>
<td>328029.91</td>
<td>54.20</td>
</tr>
<tr>
<td>0.9</td>
<td>3003.10</td>
<td>11.07</td>
<td>719380.72</td>
<td>44.81</td>
</tr>
</tbody>
</table>
4. Discussion

This study discusses the performance of several approaches to construct geometric based control charts to monitor adverse event rates. It is desired to detect the changes rapidly because the cost of an undesirable process can be quite high. A $p$ chart can be used for a process rate not quite low, such as $p = 0.1$, however, the probability-based limit $g$ chart performs better for low rates like $p = 0.01$. To detect very small positive shifts, for example for a shifted rate of $p' = p + 0.1p$, the probability limits have again more power. Other than quite small shifts, it is more powerful to use an EWMA chart which has larger in-control ARL and much smaller out-of-control ARLs. Setting in-control ARLs to observe the out-of-control ARLs also suggests that the EWMA chart needs fewer samples to detect the rate increases.
Chapter 3. Development of an Online Improvement Dashboard for Specialty Care in Veterans Affairs

1. Background

This chapter describes a baseline quality dashboard created as a demonstration for the web-based tool called Specialty Care Compass. The dashboard is intended for use by the Veterans Health Administration (VHA) as a part of the Specialty Care (SC) Collaborative. It includes practical and basic measures of access to care, care coordination, and efficiency in an accessible manner for clinical teams. This tool targets several stakeholder groups: specialist providers and teams, specialty chiefs, partner providers, and local and national leadership. Whether nation-wide or facility-wide, once the measures are defined, it is crucial to analyze individual processes to detect changes needed at the process level. To monitor progress toward achieving quality and performance improvement goals, analytical applications are used to identify if and when a change has occurred as well as the improvements needed in the system.

As emphasized in a recent article by Pronovost, “It is time for the science of health care delivery to mature and embrace systems engineering” [3]. Quality improvement techniques in health care can help provide a foundation for standardization and systematic improvement that can lower preventable and avoidable harm. In order to provide veterans with the best possible care, an assessment of the current state of VHA healthcare practices will help guide improvement planning.

The New England Veterans Engineering Resource Center (NE VERC) is one of four national Veterans Engineering Research Centers created in July 2009. Together, they represent the single largest investment in healthcare process improvement in the NE VERC partners with Northeastern University’s healthcare industrial engineering program, the Massachusetts Institute of Technology, Worcester Polytechnic Institute, and several VA centers of excellence. These partnerships develop and implement industrial
engineering solutions to improve process quality, access, safety, efficiency, and performance throughout the VA healthcare system, as well as overall New England and U.S. healthcare processes.

New England VERC is a part of the Specialty Care Collaborative planning team facilitating process experts, practitioners, and teams in sharing ideas and knowledge to learn a methodology of change. The aim is measuring progress that will build upon existing knowledge and generate new innovations to transform the delivery of Specialty Care. As part of the development of this model of care, New England VERC decided to create a web-based tool named Compass. As an introduction to this tool, we created a baseline dashboard via Excel VBA in order to monitor performance of the facilities and specialties as part of the measurement plan for teams in the collaborative. This tool is used as an introduction of control charts to clinical teams to help facilitate improved use of the data and to interpret data more accurately.

The working flow of the collaborative can be seen in Figure 13 adopted from IHI Breakthrough Series model [40]. This phase of the collaborative was comprised of 28 leadership and point of care teams representing orthopedics, urology in the surgical focus, cardiology and gastroenterology teams in the medical focus. As part of the collaborative, these teams work together to test and refine the new model of patient-centered care, focusing on access, care coordination and practice redesign. During this collaborative period, teams participate in three 2-day Virtual Learning Sessions, and maintain continual contact with each other during the intervening months between Learning Sessions (Action Periods). Additionally, three engineers from the VERC program supported the teams in the collection and interpretation of data. This process is continuous over time, and refinement is added with learning cycles which are known as “plan-do-study-act” (PDSA) cycles.
In 2001, the Institute of Medicine (IOM) defined six aims for improvement of health care in the report, "Crossing the Quality Chasm: A New Health System for the 21st Century." According to these characteristics, care should be Safe, Effective, Patient-centered, Timely, Efficient and Equitable. The Institute for Healthcare Improvement (IHI) has also defined a “Triple Aim”: 1) improving the individual experience of care; 2) improving the health of populations; and 3) reducing the per capita costs of care for populations. These dimensions have been widely adopted by the organizations to implement change and improve health care. In terms of our project, the dashboard created for the collaborative assists teams in understanding the current baseline performance and determining where changes are needed in order to deliver the highest quality healthcare to the patients. The measures included in this dashboard target are part of the IOM dimensions and Triple Aim as illustrated in Table 9.

Table 9. Dashboard measures aligned to the Triple Aim and Six Dimensions of IOM

<table>
<thead>
<tr>
<th>Triple Aim</th>
<th>IOM Dimensions</th>
<th>Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Better Patient Care</td>
<td>Effective</td>
<td>E-consult utilization</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fecal occult blood test (FOBT) tracking (clinical measures are in development)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Proportion of days covered (PDC) in DAPTposDES</td>
</tr>
<tr>
<td>Better Population Health</td>
<td>Timely</td>
<td>Percent appointments scheduled within 0-14 days</td>
</tr>
<tr>
<td>Cost Reduction</td>
<td></td>
<td>Average days to 3rd next available appointment</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Average new patient wait time</td>
</tr>
<tr>
<td></td>
<td>Efficient</td>
<td>Incomplete consults</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Missed opportunities</td>
</tr>
</tbody>
</table>
2. Content of the Compass

The Compass is an online quality tool and includes practical and basic measures of access, care coordination and efficiency in an accessible manner for specialist teams. The baseline dashboard created via Excel VBA is an introduction in order to provide data and graphs on some of the important measures for specialty collaborative teams. These measures, provided in Table 9, were pulled electronically from VA database and eventually migrated to the online Compass which is available for all users across VHA. All measures are presented graphically as either a run chart or a Statistical Process Control (SPC) chart. In this initial Excel based version, each tab has two or three measures, with graphs at the top and the data in a table at the bottom. The VA uses run charts to obtain simple visualizations of data over time, enabling clear interpretation of trends and changes. National data is also displayed where appropriate for comparison and gauging progress. There are arrows that indicate "Higher is better" or "Lower is better" for that particular measure.

Table 10. Definitions of the measures included in the dashboard

<table>
<thead>
<tr>
<th>Tabs</th>
<th>Measures &amp; Definitions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Access Core</td>
<td><strong>Average days to 3rd next available appointment</strong> - breakdown by new and established patients</td>
</tr>
<tr>
<td></td>
<td><strong>Average new patient wait time</strong> - time from appointment creation to appointment scheduled date</td>
</tr>
<tr>
<td></td>
<td><strong>E-consult utilization</strong> - number of e-consults completed in facility</td>
</tr>
<tr>
<td>Access Supporting Measures</td>
<td><strong>Percent appointments scheduled within 0-14 days</strong> - breakdown by new and established patients</td>
</tr>
<tr>
<td></td>
<td><strong>Missed opportunities</strong> - appointments cancelled on the same day of the appointment and no-shows, breakdown by appointment status (cancelled by clinic, cancelled by patient, no shows)</td>
</tr>
<tr>
<td>Care Coordination</td>
<td><strong>Octane ratio</strong> - proportion of established to new patients</td>
</tr>
<tr>
<td></td>
<td><strong>Incomplete consults</strong> - ratio of incomplete consults [cancelled + discontinued], breakdown by cancelled and discontinued</td>
</tr>
<tr>
<td>FOBT Tracking</td>
<td><strong>Fecal occult blood test (FOBT)</strong> - one of the colonoscopy screening methods, contains number of days elapsed until the colonoscopy</td>
</tr>
<tr>
<td>DAPTpostDES</td>
<td><strong>Proportion of days covered (PDC)</strong> during 1 year post procedure date of dual antiplatelet therapy (DAPT) after drug-eluting coronary stents (DES), duration of DAPT</td>
</tr>
</tbody>
</table>
The measures that are included in the tool are presented in Table 10. This tool consists of five tabs, first of which provides some information about the measures and control charts for the user. The following four tabs are Access Core Measures, Access Supporting Measures, Care Coordination and FOBT (Fecal occult blood test) tracking, used for gastroenterology teams. The tool for a cardiology team would feature a cardiology clinical measure, such as DAPTposDES (dual antiplatelet therapy after drug-eluting coronary stents). More detailed explanations about these measures are provided in the next section.

3. Quality Measures

There are two types of office visits; new patient and established patient, and some of the measures in this tool distinguish between new and established patients. A new patient is one who has not received any professional services from the physician or another physician of the same specialty working in the same group practice, within the past two years. An established patient is one who has received professional services within the past two years. Achieving the goals of the SC Collaborative translates to achievement of access, practice redesign, and care coordination goals.

3.1. Access measures

Access is a measure of the patient's ability to look for care and receive care with the provider that they have chosen, at the time they choose [41]. The goal is to achieve timely and reliable access to the SC services by offering patients single or group appointments and non-appointment options. For the collaborative, the access measure is divided into two categories: Core Measures and Supporting Measures. The Core Measures (average days to third next available appointment, average new patient wait time and e-consult utilization) are outcome measures that show progress in meeting the collaborative aims. The Supporting Measures (percent appointments scheduled within 0-14 days and missed opportunities) are additional measures that assist teams in making improvements in order to achieve the aims.
The third next available appointment is used to assess the average number of days to the third next available appointment for an office visit for each clinic and/or department. According to the IHI, in contrast to first and second available appointments that are often the result of last minute cancellations, working patients into the schedule, or other events, the third next available appointment reflects true appointment availability [42].

The average wait time in days is calculated from wait time for each appointment by counting the number of days between the appointment encounter date and the earliest of these three dates: the appointment creation date, the date the patient was entered on the Electronic Wait List (EWL) (if entered) and the date the appointment was originally created if the appointment was the result of a rescheduled clinic cancellation. New patient records are calculated from the appointment creation date whereas the established patient records are calculated from the appointment desired date.

E-consult utilization measure represents the number of e-consults completed in a facility. The aim is to increase care provided outside of single face-to-face visits (group visits, telephone care, tele-health, secure messaging, nurse visits, etc.). This type of alternate care was of particular interest to teams and the collaborative goals and was singled out for measurement in the dashboard for the time being.

The percent of appointments scheduled within 14 days is the formal measure currently in place and the goal of the collaborative is to get the patient appointments in by 7 days. This measure is broken down into new and established patients. The denominator is all new/established patient appointments while the numerator is the number of appointments from the denominator that had a wait time less than 14 days.

Missed opportunities are the appointments cancelled by clinic, cancelled by patient and no shows. In addition to a run chart that illustrates the breakdowns of missed opportunities, the percent of missed opportunities is plotted on a p chart where the values are calculated as total missed opportunities over total appointments which is given in equation (18).
3.2. Care coordination measures

Care coordination is a function that supports information-sharing across providers, patients, types and levels of service, sites, and time frames. The goal of coordination is to ensure that patients’ needs and preferences are achieved and that care is high quality and efficient.

The octane ratio measure is the number of established patients seen in the clinic divided by the number of new patients seen. The aim is to improve returns to Primary Care with transition plans, thereby creating capacity to see additional new patients.

The incomplete consults measure includes cancelled, discontinued, discontinued/edit and expired consults. The incomplete consult rates are obtained in the ratio of the cancelled and discontinued appointments to the total consults. The aim is to improve the number of appropriate patients arriving in the SC with the correct work-up. The total number of consults consists of

- Active, complete, partial results, pending, scheduled and unreleased consults which are named as Reminder in the Excel based tool.
- Cancelled, discontinued, discontinued/edit and expired consults which constitute incomplete consults.

3.3. FOBT tracking

Fecal occult blood test (FOBT) is one of the colorectal cancer (CRC) screening tests. Colonoscopy is used as a follow-up test if anything unusual is found during one of the other screening tests. The process flow for CRC screening can be seen in Figure 14. Etzioni et al states that the VHA rates for colorectal cancer screening are significantly higher than the national average. However, 41% of patients with positive
FOBTs failed to receive follow-up testing [43]. This measure aims to track follow-up rates by considering the percent of colonoscopy within 30/60/90 days, follow-up rates within 6 months, percent of positive FOBTs without colonoscopy and average days elapsed until the colonoscopy after the patient gets a positive FOBT result.

![CRC screening process](image)

**Figure 14. CRC screening process**

### 3.4. DAPTposDES tracking

This measure relates to coronary artery disease (CAD), which is listed as a condition of high prevalence and impact by the Centers for Medicare & Medicaid Services (2009). Drug-eluting stents (DES) are used in the treatment of CAD. Appropriate antiplatelet therapy is necessary after DES implantation to decrease the risk of stent thrombosis. Dual antiplatelet therapy (DAPT), the use of aspirin combined with clopidogrel for 12 months after DES placement, significantly reduces the risk of stent thrombosis. However, this is not followed in up to 25% of cases [44]. Therefore, this measure can help improve the safety and effectiveness of DES placement. PDC represents the proportion of days covered during 1 year post procedure date and is calculated as \[
\text{PDC} = \frac{\text{number of days of treatment}}{365}.
\]
4. Monitoring Methodologies

4.1. Run chart

The run chart (or time series plot) is a simple way to plot data values over time. This type of chart is widely used by the VA and implemented in the dashboard, but it should be noted that a run chart is not enough to draw a conclusion about the data and control charts also need to be used to make solid interpretations. Interpreting a run chart to identify special cause variation, if any, requires observing the plot points that differ from the process mean. A trend of multiple data points above or below the mean, a zig-zag pattern, or considerably different individual points could be a sign for special cause variation and hence, instability in the process. With a control chart, if variation in the process remains within the upper and lower control limits, the process is considered to be in control. Variation outside these limits indicates instability, requiring intervention or improvement.

4.2. \( p \) chart

The \( p \) chart is prevalent in plotting the proportion of defective items in a process per subgroup. The data collected are the number of defective items in each subgroup, which is assumed to follow a binomial distribution with a proportion parameter \( p \). The control limits for a given \( p \) and sample size \( n_i \) can be calculated with the equations in (19) [10, 15, 16]. The 3\( \sigma \) limits are commonly used to set a range for the process parameter at 0.27% control limits.

\[
\begin{align*}
UCL &= p + k \sqrt{\frac{p(1-p)}{n_i}}, \\
LCL &= p - k \sqrt{\frac{p(1-p)}{n_i}}
\end{align*}
\]  \tag{19}

When the process fraction \( p \) is unknown, it needs to be estimated from observed data. Then, for example, if there are \( D_i \) events in month \( i \) of \( m \) samples, the process fraction can be estimated as
\[
\hat{p}_i = \frac{D_i}{n}, \quad i = 1, \ldots, m
\]  

(20)

A common tool for monitoring continuous measures is a \( p \) chart for a binary variable created by comparing the continuous outcomes to a threshold chosen based on expertise in the field. This results in low performance of detection from converting a continuous variable to a discrete variable [45]. An example for using thresholds is the 14 days or 30 days thresholds in the VA measures.

4.3. Individuals chart

In the situations in which the sample size used for the process monitoring is \( n = 1 \), an individuals chart can be constructed. The individuals chart plots each observation as a separate data point. In many applications, the moving range of two successive observations is used as the basis of estimating the process variability where the moving range is

\[
MR_i = |x_i - x_{i-1}|, \quad i = 1, \ldots, n
\]  

(21)

which is the absolute value of the first difference (e.g., difference between two consecutive points) of the data [4]. One can also establish a control chart of the moving ranges (MR). The control limits for individuals and MR charts are given in Table 11. When there are individual units, these control charts are useful. However, the cumulative sum (CUSUM) and exponentially weighted moving average (EWMA) control charts are a better alternative especially if the shift in the process mean is small.

| Table 11. Control limits for individual observations and moving range (MR) |
|-----------------------------|-----------------------------|
| Upper Control Limit (UCL)   | \( \bar{x} + 3 \frac{MR}{d_2} \) | \( D_4 \bar{MR} \) |
| Center Limit (CL)           | \( \bar{x} \)               | \( MR \)           |
| Lower Control Limit (LCL)   | \( \bar{x} - 3 \frac{MR}{d_2} \) | \( D_3 \bar{MR} \) |
5. Sample Outputs from a VA Facility

The datasets for the measures are pulled by the VA electronically from their databases, then entered into the “Data” tab of Excel dashboard, and the macros are executed as captured in Figure 15. It shows where the national data for Gastrointestinal (GI) specialty care is entered which is followed by sample facility data. The charts in the other tabs are dynamically created as shown in the following figures when the user clicks the Run button.

Figure 15. Screenshot of data input tab for national data followed by facility data

Figure 16. Core appointment access measures; average days until 3rd next available appointment, e-consults and average new patient wait time
Figure 16 contains core access measures while Figure 17 has access support measures and missed opportunities. The first run chart in Figure 16 plots the average days to third next available appointments in each month indicating that average waiting days until third next appointment in the facility are higher than the national average. There might be a seasonality effect because the values have an ascending-descending pattern. One can also observe a decrease after November FY12 which is a special cause variation that needs to be investigated. The next run chart for e-consults is zero since there is no e-consult among the total consults. The third run chart indicates that the facility performs well with respect to national data but there is an increase in the average new patient wait time after December FY12.

![Appointment Access](image)

Figure 17. Support access measures; appointments scheduled within 14 days and missed opportunities

On the other hand, the run chart for percent of new patient appointments scheduled within 14 days in Figure 17 shows a better performance in recent months. The increase in the average new patient wait times in these months might be due to an outlier which needs further investigation. The percent of established patient appointments scheduled within 14 days shows a high and stable performance around 80% but the values are lower than the national data values. On the right hand side of Figure 17, the missed opportunities are monitored with a \( p \) chart and a run chart. The \( p \) chart plots the percent of missed opportunities which has an out-of-control point beyond the UCL at January FY11. One possible reason is
an increase in the appointments cancelled by clinic at this period as the run chart for breakdowns indicates. The run chart shows an increase for the breakdowns in recent months but there are almost six times the amount of appointments and missed opportunities in these months. It would be more efficient to plot percentages instead of numbers as in Figure 18, which shows the peak at January FY11 more clearly and no increase in the last months.

![Figure 18. Percent of missed opportunities instead of plotting numbers in Figure 17](image_url)

The next tab includes outputs for the care coordination measures which are octane ratio and incomplete consults as shown in Figure 19. The octane ratios of the facility, which is the ratio of number of established patients seen to the number of new patients seen, seem to be higher than the national ratios as the run chart on the left of Figure 19 indicates. However, the ratios are more stable and have decreased in the last six months. On the right side, there is a $p$ chart for the incomplete consult rates and a run chart for breakdowns. According to the $p$ chart, the incomplete consult rates of the facility are in-control although they are higher than the national rates.
Figure 19. Care coordination measures; octane ratio and incomplete consults

The last set of outputs show examples of the clinical measures that are under development, FOBT tracking and cardiovascular medicine. The first part of FOBT tracking in Figure 21 displays the percent of colonoscopy within 30/60/90 days with a cumulative run chart and bar chart. These charts illustrate similar things but both are included to make the process understandable for the nontechnical audience. The run charts in the second part of Figure 21 exhibit the follow-up rates within 6 months and average days to follow-up, respectively. The average days are calculated by dividing the total days until colonoscopy by total the number of patients in that month. The averages are as high as two months at the beginning followed by a steady decrease. The last chart is a $p$ chart for the percent of positive FOBTs without colonoscopy. The proportions for each month are estimated as the number of Nulls (positive FOBT cases without colonoscopy) over total number of patients (including Nulls). Although average days to follow-up seem to be improving, there is an increase in the percent of positive FOBTs without colonoscopy and there is an out-of-control point at the last month which is February 2012. The cause of this rise needs to be investigated and action should be taken.
The charts in Figure 21 are about dual antiplatelet therapy (DAPT) after drug-eluting stent (DES) implantation. They monitor the percent of days covered by the VA during the 1 year post procedure date (PDC) which is calculated as the number of days of treatment over 365. In the run chart, which plots the maximum, average and minimum PDCs, the PDCs seem to be improving but it is better monitored in the individuals chart. Although there are some out-of-control points between February 2009 and April 2009, the percentage is increasing afterwards.

Figure 21. DAPTpostDES measures
6. Discussion

VHA had been running the Specialty Care Collaborative which brings experts and practitioners together to develop new approaches to care, and has designed measures to understand the current performance, identify the baselines, and monitor progress. The SC Compass brings together a series of metrics into a dashboard that reflects the dimensions and principles of the PACT and SC team working in a medical home neighborhood. It includes practical and basic measures of access, care coordination and efficiency in an accessible manner for specialist teams and control charts to monitor the performances. In improvement efforts, knowing the baseline and trending of critical indicators is one of the key means to knowing if changes are taking effect. This paper consists of the measures and methodologies to monitor them, and the baseline tool created via Excel VBA which eventually migrated to the online Compass. The system was also piloted with data to illustrate the use and utility of the dashboard, and outputs are presented by New England VERC team in virtual training sessions during the Collaborative.
Chapter 4. Statistical Process Control Methods for Grouped Data in Normal and Weibull Models

1. Background

Descriptive statistics are easily computed when data are available with high precision. In various fields such as engineering and medicine, the measurements of a statistical experiment are classified into intervals rather than obtaining the exact values. The problem occurs due to using grouped data to make inferences about the parameter $\theta$ since there is no complete information about the sample. This chapter proposes several methods for monitoring a process for grouped data and compares the relative performances of these methods for detecting changes in the process mean and standard deviation. The methods discussed are midpoint approach, $p$ chart, chi-square control chart, maximum likelihood estimation (MLE) and the method of minimum chi-square. The sensitivity analyses are conducted in terms of different sample sizes, interval widths, and type of the distribution via Monte Carlo simulation. The average run lengths (ARL) are used as criterion to investigate the performances of given methods. The methods are then illustrated with a dataset from Veterans Affairs’ (VA) colonoscopy services. There are several VA quality measures in frequency data format without sample mean and variance information which makes Shewhart charts impractical. The proposed methods can be used to monitor for this type of data sets within the VA.

Grouped data is the result of observing continuous variables up to the nearest interval. There are several reasons that data are grouped: for instance, to summarize with graphical and tabular presentations [46]. In some situations, grouping is related to the confidentiality of individual records. Another situation is grouping the continuous variables into categories during data collection or publication which is the case in personal income data, where only frequencies within specified classes are available. The problem arises
due to using grouped data to draw inferences about the population parameter because the complete information about the sample is not available.

There are several studies that focus on grouped data inference. Kulldorff (1961) provides a thorough study of grouped and partially grouped samples including MLE [47-52] and midpoint approach [53]. Gjeddebaek (1968) discusses mainly from a likelihood perspective, and Haitovsky (1982) presents a comprehensive summary of methodologies [54, 55]. Heitjan (1989) provides an integrated review of the methods in grouped data inference [46].

The choice of estimation method for a grouped data affects the performance of the control chart that is used to monitor the process. The control charts that are applicable in the general multiple group cases are designed by Steiner, Geyer and Wesolowsky (1994), and later are extended to Shewhart type control charts. They propose a methodology for the design of sequential methods and expressions for the OCs of Wald sequential probability ratio tests and for the ARL of cumulative sum (CUSUM) [56]. Steiner also suggests a version of exponentially weighted moving average (EWMA) control charts to monitor grouped data for process shifts and compares with CUSUM charts [57]. This study follows a similar manner but considers nonsymmetric distributions in addition to symmetric distributions; it also includes sensitivity analyses for additional methods such as MLE, chi-square control chart and minimum chi-square estimation method.

Control charts are used to detect assignable causes that affect process stability. It is standard practice to use an $\bar{x}$ chart to detect the changes that shift the process mean. To obtain an $\bar{x}$ and $s$ chart for the data having sample size $n$ and $m$ subgroups as in Table 12, if no standard is given for the population variance, the mean and variance are estimated as
\[
\bar{x}_i = \frac{x_{11} + x_{12} + x_{13} + \cdots + x_{1n}}{n}, \quad s_i^2 = \frac{\sum_{i=1}^{n} x_i^2 - n \bar{x}^2}{n-1}
\]

(22)

for each month \(i = 1, \ldots, m\). However, the \(\bar{x}\) and/or \(s\) values or the actual observations and subgroup sizes are usually not available to estimate the parameters. Instead, the continuous observations are classified into groups which is the case in some of the VA data sets, represented with quantiles as shown in Table 12.

Table 12. An example for grouped data; patient appointment waits in a facility

<table>
<thead>
<tr>
<th>Month</th>
<th>0-30 days</th>
<th>31-60 days</th>
<th>61-90 days</th>
<th>91-120 days</th>
<th>&gt;120 days</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(Q_{11})</td>
<td>(Q_{12})</td>
<td>(Q_{13})</td>
<td>(Q_{14})</td>
<td>(Q_{15})</td>
</tr>
<tr>
<td>2</td>
<td>(Q_{21})</td>
<td>(Q_{22})</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>\ldots</td>
<td>\ldots</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(m)</td>
<td>(Q_{m1})</td>
<td>(Q_{m2})</td>
<td>\ldots</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Number of data points (frequency) | \(Q_1\) | \(Q_2\) | \(Q_3\) | \(Q_4\) | \(Q_5\) |

Multiple grouped data is practically held with an ad hoc method that uses the midpoints or endpoints of the intervals as a representative value. Most of the control charts that are used to monitor grouped data are based on assigning either a group endpoint or midpoint to each observation. However, this method has some disadvantages: it would create a bias when estimating the process mean unless the distribution is symmetric within each interval. In addition, the last interval could be defined as infinite. For the purpose of finding the best method to monitor frequency data efficiently, chi-square control chart, MLE and minimum chi-square estimation methods are also used. Monte Carlo simulations are built for each method to observe and compare their performances. The criterion used to evaluate the performance of the charts is the ARL indicating how many samples are expected to be taken in order to detect an out-of-control signal. For an in-control process, ARL should be large corresponding to a low false alarm rate. In contrast, an out-of-control process would have a low ARL implying that the change is detected quickly. These methods are illustrated and tested using actual data from the colonoscopy services of the VA.
2. Methodologies for Monitoring Grouped Data

2.1. The midpoint approach – Quantile chart

Assume that the data values are unknown and they are estimated with the midpoint of the interval to which corresponding data value belongs. That is, the values in the first interval will be estimated as \( m_1 \) which is the midpoint of the first interval. An approximation for \( \mu \) are then obtained as

\[
\hat{x}_i \approx \frac{\sum_{j=1}^{k} m_j Q_j}{\sum_{j=1}^{k} Q_j}
\]

where \( m_j \) is the midpoint of \( j^{th} \) quantile \((j = 1, \ldots, k)\) and \( Q_j \) is the number of data points in \( j^{th} \) quantile.

The approximation for the standard deviation is given in equation (24).

\[
s_i \approx \sqrt{\frac{\sum_{j=1}^{k} m_j^2 Q_j - \left[ \frac{\sum_{j=1}^{k} m_j Q_j}{\sum_{j=1}^{k} Q_j} \right]^2}{\sum_{j=1}^{k} Q_j - 1}}
\]

Without the raw data, the statistics, such as the mean and variance, cannot be calculated. However, these statistics can be approximated. This example illustrates approximation of the mean and variance through the midpoint approach.

Suppose that \( m = 100 \) samples with size \( n = 15 \) in a process that is grouped into \( k = 4 \) classes. The data is obtained by generating random numbers from a Normal distribution with \( \mu = 70 \) and \( \sigma = 15 \). It is assumed that there is a continuous measurement that is unobservable and the grouped data that consist \( k = 4 \) classes with interval size of 30 are available. The resulting grouped data is given partially in Table 13 where each cell represents the frequency of the observations falling into that interval.

The sample mean estimated from the actual data is 69.691 and the midpoint approach is estimated \( \bar{x} \) as 70.218 from the grouped data. The sample mean that is estimated by midpoint approach is close to the
classic estimation of the true mean $\mu = 70$. The sample standard deviation is estimated by classic and midpoint approach as 15.201 and 17.157. The sample mean is an unbiased estimate of the process mean for this symmetric dataset, but the process variance is overestimated by the midpoint method [57].

Table 13. Grouped data with $\bar{x} = 69.691$ and $\bar{s} = 17.157$ by midpoint approach

<table>
<thead>
<tr>
<th>$i$</th>
<th>Intervals</th>
<th>Midpoint Approach</th>
<th>Actual Data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0-30</td>
<td>$\bar{x}_i$</td>
<td>$s_i$</td>
</tr>
<tr>
<td>1</td>
<td>0-30</td>
<td>31</td>
<td>8</td>
</tr>
<tr>
<td>2</td>
<td>0-30</td>
<td>26</td>
<td>7</td>
</tr>
<tr>
<td>3</td>
<td>0-30</td>
<td>21</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>1-30</td>
<td>27</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>0-30</td>
<td>31</td>
<td>9</td>
</tr>
<tr>
<td>6</td>
<td>0-30</td>
<td>18</td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td>2-30</td>
<td>20</td>
<td>12</td>
</tr>
<tr>
<td>8</td>
<td>2-30</td>
<td>29</td>
<td>5</td>
</tr>
<tr>
<td>9</td>
<td>0-30</td>
<td>34</td>
<td>6</td>
</tr>
<tr>
<td>10</td>
<td>0-30</td>
<td>21</td>
<td>13</td>
</tr>
<tr>
<td>11</td>
<td>0-30</td>
<td>21</td>
<td>6</td>
</tr>
<tr>
<td>12</td>
<td>0-30</td>
<td>25</td>
<td>7</td>
</tr>
<tr>
<td>13</td>
<td>1-30</td>
<td>29</td>
<td>11</td>
</tr>
<tr>
<td>14</td>
<td>1-30</td>
<td>27</td>
<td>9</td>
</tr>
<tr>
<td>15</td>
<td>0-30</td>
<td>20</td>
<td>7</td>
</tr>
</tbody>
</table>

Figure 22(a) plots the variables-based and grouped data sample means with Shewhart chart control limits. The estimated means and control limits are quite close. This method is an ad hoc solution when the continuous observations are not available, but some shortcomings for this approach are present.

Figure 22. (a) $\bar{x}$ chart with midpoint approach, (b) data in the form of a frequency distribution
Although the midpoint approach is easy to apply, the estimations might be biased because the midpoints are used instead of the actual data values which may not truly represent the actual values. Using the midpoints when compared to the endpoints of the intervals results in less bias, but the bias still exists unless the distribution is uniform [58]. If the data points plotted at the left and right side of the midpoint are considered as positive and negative values as in Figure 22(b), they would cancel out for the uniform distribution case. The same cancellation can be considered for left and right tails of a symmetric distribution. If the data set is not normal, it can be transformed to normally distributed data by removing the skewness in the data and by getting the skewness at zero. However, since this is mostly inapplicable in reality, the lack of uniformness or nonsymmetric shape would result in bias. In this case, it would be more accurate to use a theoretical mean instead of approximating with \( m_i \) since this approximation is not always accurate, especially if the underlying distribution is not symmetric. The Weibull distribution, which is widely used as a fit to life time data, is included in the study to investigate the effect of the type of the distribution.

The midpoint approach has further difficulty if the first or last interval has no lower or upper bound. This is not the case in the example above; however, infinite intervals are also taken into account in the simulations. In order to determine the midpoints, the definition in equation (25) of Steiner is applied where \( k \)-interval endpoints denoted by \( t_j, j = 1, ..., k \) [57]. In this definition, the end groups are assigned scores based on the most extreme limits and the distance to the second most extreme scores on either side. For the groups with equal width, this approach is reasonable; however, other definitions of the weights are possible.

\[
m_j = \begin{cases} 
\frac{3t_1 - t_2}{2} & \text{for } j = 1 \\
\frac{t_{j-1} + t_j}{2} & \text{for } 2 \leq j \leq k - 1 \\
\frac{3t_{k-1} - t_{k-2}}{2} & \text{for } j = k
\end{cases}
\] (25)
2.2. Maximum likelihood estimation

The first part of the simulations consists of the performances of the control charts when the process mean and standard deviation are estimated by the classical method and midpoint approach. This follows the estimation of the parameters by maximizing likelihood functions; this is expected to be a more accurate theoretical approach for parameter estimation and consequently better for detection. In the process of maximizing the likelihood function of grouped data, there is no unified method for obtaining ML estimates of the distribution parameters, except for maximizing numerically [59]. For a grouped sample, the likelihood function of the parameter of interest $\theta$ with sample of size $n$ is

$$L(\theta) = \prod_{j=1}^{k} \left( \int_{t_{j-1}}^{t_j} f(x) dx \right)^{Q_j}$$

(26)

where $k$-interval endpoints denoted by $t_j$, $j = 1, ..., k$ and $t_k = \infty$ [57]. $Q_j$ is the number of points falling into the interval $I_j = (t_{j-1}, t_j]$ and $\sum_{j=1}^{k} Q_j = n$. The general notations are also given in Table 14.

<table>
<thead>
<tr>
<th>$j$</th>
<th>Class interval</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$(0, t_1]$</td>
<td>$Q_1$</td>
</tr>
<tr>
<td>2</td>
<td>$(t_1, t_2]$</td>
<td>$Q_2$</td>
</tr>
<tr>
<td>$\vdots$</td>
<td>$\vdots$</td>
<td>$\vdots$</td>
</tr>
<tr>
<td>$k-1$</td>
<td>$(t_{k-2}, t_{k-1}]$</td>
<td>$Q_{k-1}$</td>
</tr>
<tr>
<td>$k$</td>
<td>$(t_{k-1}, \infty)$</td>
<td>$Q_k$</td>
</tr>
</tbody>
</table>

The likelihood function for grouped data is written in equation (27) when the actual data is distributed normally. Note that the probability for the last interval is considered separately due to infinity.

$$L(\mu, \sigma) = P(t_0 < X < t_1)^{Q_1} P(t_1 < X < t_2)^{Q_2} \cdots P(t_{k-2} < X < t_{k-1})^{Q_{k-1}} P(t_{k-1} < X < \infty)^{Q_k}$$

$$= \left( \prod_{j=1}^{k-1} \left[ F(t_j) - F(t_{j-1}) \right]^{Q_j} \right) \left[ 1 - P(X < t_{k-1}) \right]^{Q_k}$$

(27)
For the three-parameter Weibull distribution with the cdf $F(x) = 1 - e^{-\left(\frac{x - \gamma}{\lambda}\right)^{\beta}}$ where $\lambda$ is the scale parameter, $\beta$ is the shape parameter, and $\gamma$ is the location parameter, the log-likelihood function is given in equation (28).

$$\ln L(\lambda, \beta, \gamma) = \sum_{j=1}^{k-1} Q_j \ln \left[ e^{-\left(\frac{t_j - \gamma}{\lambda}\right)^{\beta}} - e^{-\left(\frac{t_{j-1} - \gamma}{\lambda}\right)^{\beta}} \right] + Q_k \ln \left[ 1 - e^{-\left(\frac{t_k - \gamma}{\lambda}\right)^{\beta}} \right] \tag{28}$$

The next step in maximizing the log-likelihood function is to take the derivative with respect to the parameters. The first derivative with respect to $\lambda$ is given in equation (29). The first and second derivatives for Normal with respect to $\mu$ and the second derivative for Weibull distribution with respect to $\lambda$ are given in the Appendix. There are no closed form solutions for these equations, thus they need to be solved by iterative numerical techniques.

$$\frac{\partial \ln L(\lambda, \beta, \gamma)}{\partial \lambda} = \sum_{j=1}^{k-1} Q_j \left[ \frac{(t_j - \gamma)\beta e^{-\left(\frac{t_j - \gamma}{\lambda}\right)^{\beta}} \left(\frac{t_j - \gamma}{\lambda}\right)^{\beta-1}}{\lambda^2} - \frac{\beta(t_j - \gamma) e^{-\left(\frac{t_j - \gamma}{\lambda}\right)^{\beta}} \left(\frac{t_j - \gamma}{\lambda}\right)^{\beta-1}}{\lambda^2} \right]$$

$$- Q_k \frac{(t_{k-1} - \gamma)\beta e^{-\left(\frac{t_{k-1} - \gamma}{\lambda}\right)^{\beta}} \left(\frac{t_{k-1} - \gamma}{\lambda}\right)^{\beta-1}}{\left(1 - e^{-\left(\frac{t_{k-1} - \gamma}{\lambda}\right)^{\beta}}\right)\lambda^2} \tag{29}$$

In this study, the Newton Raphson method is used in order to obtain the values of the parameters which maximize the likelihood function. This maximization process iterates equation (30) where $\theta_0$ is a starting point for the algorithm and $\theta_1, \theta_2, ...$ are the successive estimates. The $f$ is the function to be maximized which is the log-likelihood function, and $f'$ and $f''$ represents first and second derivatives of log-likelihood function respectively. The $\theta_i$ that gives the maximum log-likelihood value will be taken as the approximate ML estimation.

$$\theta_{i+1} = \theta_i - \frac{f'(\theta_i)}{f''(\theta_i)} \tag{30}$$
2.3. \( p \) chart

The \( p \) chart is a popular choice to plot the proportion of defective items in a process per subgroup. The data that is collected is the number of defective items in each subgroup, which is assumed to follow a binomial distribution with a proportion parameter \( (p) \). The control limits for a given \( p \) and sample size \( n_i \) can be easily calculated with the equations in (31) [15].

\[
\begin{align*}
UCL &= p + k \sqrt{\frac{p(1-p)}{n_i}}, \\
LCL &= p - k \sqrt{\frac{p(1-p)}{n_i}}
\end{align*}
\]  

(31)

In health care, a common approach for monitoring continuous measures is using a \( p \) chart for a binary variable. This is created by comparing the continuous outcome measures to a threshold value that is chosen based on expertise in the field. This results in a low performance of detection because of converting a continuous variable to a discrete variable. This method is also included in the simulation study by setting a threshold.

2.4. Chi-square control chart

The simultaneous control of two or more variables is usually referred as multivariate quality control. After Shewhart suggested the use of the chi-square test as a fifth criterion to detect the changes in a process in 1931, modeling with the chi-square control chart has been widely studied. The basic structure of this chart was laid out originally by Duncan (1950), Marcucci (1985), Nelson (1987), and Montgomery (2005) [4, 25-27]. This type of control chart is also called as generalized \( p \) chart [25].

When the output from a process has several categories, the process may be monitored in two ways. The first is construction of individual \( p \) charts of each category and the second is designing a single chi-square chart. Studies conclude that these two methods are statistically equivalent [26]. One might prefer to use a chi-square chart since it presents a single summary of control rather than multiple charting. The chi-square table that is used to plot the chart can be used for finding the causes of a shift in the process. For
example, if a large chi-square statistic is the result of a large contribution of a particular component, it reflects out-of-control process due to that component. Note that the interpretations of a $p$ chart and a chi-square chart are different. A low level shift in a $p$ chart indicates an improving trend that is no longer random or the reflection of a calculation error in the chi-square chart.

The chi-square control chart plots the statistic obtained from a goodness-of-fit test that is applied to each group. The statistic $\chi^2$ which is proposed by Pearson (1900) is given as

$$
\chi_i^2 = \sum_{j=1}^{k} \frac{(O_j - E_j)^2}{E_j}
$$

for $i = 1, \ldots, m$ and $j = 1, \ldots, k$ for $m$ subgroups and $k$ categories. The $O_j$ values represent observed values which are the number of observations falling into each interval ($Q_j$) in the grouped data. The expected values ($E_j$) are calculated using the probability of falling into the $k^\text{th}$ interval. The probability based upper and lower control limits are given in Table 15 as well as $3\sigma$ control limits. Although $3\sigma$ limits are preferred due to its applicability, the control chart with probability limits is expected to perform better.

Table 15. Control limits for chi-square chart

<table>
<thead>
<tr>
<th></th>
<th>3σ limits</th>
<th>Probability limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>UCL</td>
<td>$(k - 1) + 3\sqrt{2(k - 1)}$</td>
<td>$\chi^2_{(1-\alpha/2,k)}$</td>
</tr>
<tr>
<td>LCL</td>
<td>$(k - 1) - 3\sqrt{2(k - 1)}$</td>
<td>$\chi^2_{(\alpha/2,k)}$</td>
</tr>
</tbody>
</table>

2.5. **The method of minimum chi-square**

The minimum chi-square method of estimation is applicable to situations when the data is in the form of frequencies [60]. The method is an extension of the chi-square goodness-of-fit test and has asymptotic properties similar to ML. For $k$ classes ($j = 1, \ldots, k$), the probability of falling in the $j^\text{th}$ class is defined as:

$$
p_j(\theta) = \int f(x; \theta) \, dx
$$

(33)
This is calculated with the cdf of the underlying distribution (Normal or Weibull) in the Monte Carlo simulation studies. Also, \( \hat{p}_i \) is the proportion of the observed values falling into \( j^{th} \) class. The minimum \( \chi^2 \) estimate of \( \theta \) is the value of \( \theta \) which minimizes equation (34). This \( \theta \) would make the \( \chi^2 \) goodness-of-fit test statistic as small as possible. A possible approach in order to find the \( \theta \) that minimizes the \( \chi^2 \) is searching through an interval. The simulation in this study calculates \( \chi^2 \) values for each \( \theta \) within an interval, and finds the \( \theta \) with minimum \( \chi^2 \) value.

\[
\chi_i^2 = \sum_{j=1}^{k} \frac{[\hat{p}_i - p_i(\theta)]^2}{p_i(\theta)}
\]  (34)

This method is used in the simulation studies as an alternative to the ML estimation. Berkson questions the sovereignty of the MLE in his paper and also argues that minimum chi-square is the primary principle of estimation [61]. The method is much easier to apply than MLE which requires vast amount of derivatives and unsolvable equations.

2.6. Simulation with theoretical approaches: MLE and minimum chi-square method

The datasets used in this section are generated from Weibull(\( \lambda = 50 \) and \( \beta = 3 \)). First, it is intended to estimate distribution parameters with the MLE so that mean, variance, and control limits are calculated using these estimations. Since the equations are impossible to solve simultaneously in the classic MLE procedure, the Newton Raphson method is used to solve iteratively in MATLAB. The flow chart of this simulation is presented in Figure 23. The first and second derivatives in equation (29) and appendix (1) and \( \lambda_i \) are used to calculate \( \lambda_{i+1} \). The \( \lambda \) with maximum log-likelihood function is recorded as the MLE of the scale parameter.
Figure 23. Flow chart of Monte Carlo simulation with Newton Raphson method to estimate $\lambda$ of Weibull distribution

Although this method gives an estimation of $\lambda$ very close to the actual value of 50 in some of the iterations, the method is not always converging especially if the starting point is not close to the actual value. This iterative approach also requires known $\beta$ and $\gamma$ in order to estimate $\lambda$. Although Excel Solver is an easier option, it gives a local optimum which is very close to actual value. Table 16 shows the inputs and the local optimum $\lambda$. The equations that are needed to solve for the MLE of Normal distribution are even more complex than the Weibull equations, therefore we decided to apply the minimum chi-square estimation. For a grouped sample of (18, 69, 13, 0, 0) with 5 intervals, the method estimates $\lambda = 48.55$ and $\beta = 3.37$ which is very close to the actual values $\lambda = 50$ and $\beta = 3$ which are used to generate random samples.

### Table 16. Estimating $\lambda$ with Excel Solver

<table>
<thead>
<tr>
<th>interval</th>
<th>$Q$</th>
<th>$t$</th>
<th>OBJ (lnL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>18</td>
<td>0</td>
<td>-83.7195</td>
</tr>
<tr>
<td>2</td>
<td>69</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>13</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>90</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>120</td>
<td></td>
</tr>
</tbody>
</table>

3. Simulation Studies

In the statistical process control, there are two distinct phases of control charting. In Phase I, a set of process data is gathered and analyzed to test retrospectively whether the process was in control when the first subgroups are obtained. Trial control limits are used in this phase. The objective of the Phase I control charts is to obtain an in-control set of observations so that control limits can be set for Phase II. In Phase II, the control charts are used for testing whether the process remains in-control for the future products. The simulations in this article consist of simultaneous monitoring of $k$ classes in Phase II.
Monte Carlo simulation studies demonstrate the application of the methods for detecting change. The flow chart of the simulation is presented in Figure 24. Different methods are used to monitor the processes for the generated and grouped datasets. In review of these methods, they are:

- Classic $\bar{x}$ and $s$ chart for the average number of days until an event,
- Quantile chart (midpoint approach) for the average number of days until an event,
- Maximum likelihood estimation for the distribution parameters,
- $p$ chart for the percent of an event within $T$ (threshold) days,
- Chi-square chart for the number of days until an event for each interval (both probability and $3\sigma$ control limits) and
- Minimum chi-square estimation method to estimate the distribution parameters.

Figure 24. Flow chart of Monte Carlo simulation to monitor process mean
3.1. Detecting the shift in the process mean

The simulation results are presented in Table 17 for both Normal($\mu = 80, \sigma = 25$) and Weibull($\lambda = 55, \beta = 2, \gamma$) distributions. The expected value for this Weibull distribution is 48.746 while the standard deviation is 25.479. Case 1 is the baseline simulation when the sample size $n$ is 100 and width of the grouping interval $w$ is 30. The $\delta$ in normal table and $\gamma$ in Weibull table denote the positive shifts in the process mean. When $\delta$ and $\gamma$ are 0, the process is in-control and the given ARL represents the in-control ARL ($ARL_0$).

![Normal ARL curves for case 1](image)

Normal, $n = 100, w = 30$, the shifts $\delta = 0.1, 0.2, 0.3, 0.4$ approximately corresponds to $0.1\sigma, 0.2\sigma, 0.3\sigma, 0.4\sigma$ respectively

![Weibull ARL curves for case 1](image)

Weibull, $n = 100, w = 30$, the shifts with $\gamma = 0.1, 0.2, 0.3, 0.4$ approximately corresponds to $0.1\sigma, 0.2\sigma, 0.3\sigma, 0.4\sigma$ respectively

Figure 25. ARL curves for case 1
When a normal process is in control, $\bar{x}$ chart has approximately $ARL_0 = 370$. Therefore, a close performance to the actual method by giving $ARL_0$ close to 370 is expected for the methods used to monitor grouped data. The resulting ARL curves for case 1 are given in Figure 25: the x-axis shows the positive shifts in the process mean while the y-axis shows the average run lengths until that shift is detected. The quantile chart has a slightly better performance than the probability limit chi-square chart and also detects the shifts faster. Note that the $ARL_0$ for the chi-square chart is pretty low, especially for three-sigma control limits. Chi-square tests become undependable if some expected values are low. As a rule of thumb, if any expected value is 0 or more than one-fifth of the expected values are less than five, the chi-square test is unreliable. Based on these initial simulation results, it is investigated how to make the control charts perform better with different sample sizes and interval widths as in Table 17.

![Graph](image_url)

Normal, $n = 500$, $w = 10$, the shifts $\delta = 0.1, 0.2, 0.3, 0.4$ approximately corresponds to $0.1\sigma$, $0.2\sigma, 0.3\sigma, 0.4\sigma$ respectively

Weibull, $n = 500$, $w = 10$, the shifts with $\gamma = 0.1, 0.2, 0.3, 0.4$ approximately corresponds to $0.1\sigma, 0.2\sigma, 0.3\sigma, 0.4\sigma$ respectively

Figure 26. ARL curves for case 4
Table 17. Average run lengths for each method with different settings to detect shifts in the process mean

<table>
<thead>
<tr>
<th>Cases</th>
<th>$\delta$</th>
<th>Normal($\mu = 80, \sigma = 25$)</th>
<th></th>
<th>Weibull($\lambda = 55, \beta = 2, \gamma$)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$n = 100, w = 30$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>368.3484</td>
<td>304.4222</td>
<td>295.361</td>
<td>159.9422</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>43.4802</td>
<td>36.0764</td>
<td>131.594</td>
<td>27.4904</td>
</tr>
<tr>
<td></td>
<td>0.2</td>
<td>6.1606</td>
<td>6.0266</td>
<td>23.94</td>
<td>6.9864</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>2.0162</td>
<td>2.0738</td>
<td>5.6542</td>
<td>2.4936</td>
</tr>
<tr>
<td></td>
<td>0.4</td>
<td>1.1976</td>
<td>1.2326</td>
<td>2.0424</td>
<td>1.3696</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>1.0248</td>
<td>1.038</td>
<td>1.2404</td>
<td>1.0698</td>
</tr>
<tr>
<td>$n = 500, w = 30$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>368.2752</td>
<td>319.2816</td>
<td>377.5214</td>
<td>169.1642</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>4.4538</td>
<td>4.6102</td>
<td>18.9312</td>
<td>5.3248</td>
</tr>
<tr>
<td></td>
<td>0.2</td>
<td>1.0792</td>
<td>1.105</td>
<td>1.5978</td>
<td>1.186</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>1</td>
<td>1</td>
<td>1.0082</td>
<td>1.0006</td>
</tr>
<tr>
<td></td>
<td>0.4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$n = 1000, w = 10$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>369.9056</td>
<td>405.2804</td>
<td>229.98</td>
<td>188.3512</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>44.7722</td>
<td>80.1742</td>
<td>164.0242</td>
<td>53.1838</td>
</tr>
<tr>
<td></td>
<td>0.2</td>
<td>6.3242</td>
<td>9.3636</td>
<td>63.944</td>
<td>20.7788</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>2.0102</td>
<td>2.558</td>
<td>13.6956</td>
<td>5.9378</td>
</tr>
<tr>
<td></td>
<td>0.4</td>
<td>1.1874</td>
<td>1.3032</td>
<td>3.8922</td>
<td>2.2504</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>1.0212</td>
<td>1.0426</td>
<td>1.712</td>
<td>1.3172</td>
</tr>
<tr>
<td>$n = 5000, w = 10$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>369.01</td>
<td>358.4694</td>
<td>288.9314</td>
<td>201.7618</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>4.6032</td>
<td>8.1988</td>
<td>51.2446</td>
<td>15.1088</td>
</tr>
<tr>
<td></td>
<td>0.2</td>
<td>1.0754</td>
<td>1.1714</td>
<td>2.5146</td>
<td>1.6232</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>1.0002</td>
<td>1.0006</td>
<td>1.0434</td>
<td>1.0114</td>
</tr>
<tr>
<td></td>
<td>0.4</td>
<td>1</td>
<td>1</td>
<td>1.0002</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

* $n$: sample size, $w$: interval width, $\delta$ and $\gamma$: positive shifts in the process mean.
An increase in the sample size from 100 to 500 (case 2, Normal distribution) results in an $ARL_0 = 377.5$ for a chi-square with probability limits. This is the best result for this chart because expected values increase as the samples size increases. In addition, the chi-square chart performs better with a small number of intervals (wider intervals) since expected values are decreasing as the number of intervals increases.

The $ARL_0$ from quantile chart is closer to the classical method for case 3 and 4 (Normal distribution) due to narrower intervals. The shifts are detected faster when both sample size and number of intervals are increased (case 4). Figure 26 displays the ARL curves for case 4 and also shows that all the methods are detecting the changes faster for larger sample sizes. The simulations point out that quantile chart can be used as a substitute to the classical $\bar{x}$ chart when normally distributed data is grouped. The quantile chart is a good option when the interval width is $0.4\sigma$ ($w = 10$); the chi-square chart with the probability limits, rather than three-sigma limits can be constructed when the width is $1.2\sigma$ ($w = 30$) and the sample size is large enough.

Furthermore, the $ARL_0$ from the $p$ chart is high because the threshold is chosen at 60 which is close to the actual process mean of $\mu = 80$. However, it detects the changes faster when the sample size is increased. Its detection length is very high for case 1 and 3. A different threshold, like 30, yields extremely low $ARL_0 = 46.57$ as expected. In addition to the Normal distribution, Weibull distribution is also used in the simulations to observe the control chart performances for non-normal data. As seen in Table 17, $ARL_0$ from all methods are close to the ARL with Normal distribution, but Normal cases show faster detection of the shifts. Figure 27 also displays the in-control ARLs, and the ARLs when the shift parameter is $\gamma = 0.1$. 
3.2. Detecting the shift in the process standard deviation

In order to observe and evaluate the variation of a process over time, and against control limits, the \( s \) chart is plotted in conjunction with the \( \bar{x} \) chart. The \( s \) chart plots the standard deviation of each sampled subgroups. The same simulation methods that are used for the process mean are carried out in the process variance in order to study the performances of the control charts, and for detecting positive shifts in the standard deviation. The \( s \) values for the midpoint method are calculated with equation (24) while the classic method is applied as equation (22). Table 18 presents the ARLs of each method for different cases.

The performance of the midpoint approach is pretty low unlike monitoring the process mean. However, note that it has satisfactory in-control ARLs when the interval width is smaller, and it detects the shifts as fast as the classic \( s \) chart. Increasing the sample size does not improve the performance unless the interval width is small. Again the chi-square chart displays low performance similar to the process mean simulations.

Figure 27. ARL values for different interval widths (0.2\( \sigma \), 0.4\( \sigma \), 0.6\( \sigma \), 0.8\( \sigma \), 0.9\( \sigma \), 1.2\( \sigma \)) are \( w = 5, 10, 15, 20, 24, 30 \) respectively) when (a) the process is in-control and (b) the shift parameter is \( \gamma = 0.1 \)
Table 18. ARLs for each method with different settings to detect shift in Normal process variance

<table>
<thead>
<tr>
<th>Cases</th>
<th>δ</th>
<th>Classic</th>
<th>Midpoint</th>
<th>Chisq with prob limits</th>
<th>Chisq with 3σ limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>n = 100, w = 30</td>
<td>0.05</td>
<td>369.3745</td>
<td>196.604</td>
<td>287.1385</td>
<td>91.2</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>61.7323</td>
<td>101.1657</td>
<td>98.5283</td>
<td>24.1065</td>
</tr>
<tr>
<td></td>
<td>0.15</td>
<td>13.173</td>
<td>6.967</td>
<td>31.9625</td>
<td>10.268</td>
</tr>
<tr>
<td></td>
<td>0.2</td>
<td>4.6067</td>
<td>3.3305</td>
<td>12.2248</td>
<td>5.0085</td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>2.2982</td>
<td>1.9458</td>
<td>5.7645</td>
<td>2.897</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.5133</td>
<td>1.42</td>
<td>3.3887</td>
<td>1.936</td>
</tr>
<tr>
<td>n = 500, w = 30</td>
<td>0.05</td>
<td>368.2987</td>
<td>112.1416</td>
<td>80.9237</td>
<td>68.0316</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>11.0716</td>
<td>34.7352</td>
<td>42.2041</td>
<td>10.6438</td>
</tr>
<tr>
<td></td>
<td>0.15</td>
<td>1.82</td>
<td>8.9648</td>
<td>5.4509</td>
<td>2.5288</td>
</tr>
<tr>
<td></td>
<td>0.2</td>
<td>1.0659</td>
<td>3.5952</td>
<td>1.7178</td>
<td>1.245</td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>1.0013</td>
<td>1.989</td>
<td>1.1275</td>
<td>1.0328</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>1.4004</td>
<td>1.0122</td>
<td>1.0019</td>
</tr>
<tr>
<td>n = 100, w = 10</td>
<td>0.05</td>
<td>368.1282</td>
<td>456.6593</td>
<td>221.8157</td>
<td>77.5418</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>59.7915</td>
<td>296.9203</td>
<td>72.8780</td>
<td>21.7395</td>
</tr>
<tr>
<td></td>
<td>0.15</td>
<td>13.1058</td>
<td>95.4082</td>
<td>21.8242</td>
<td>10.2033</td>
</tr>
<tr>
<td></td>
<td>0.2</td>
<td>4.5945</td>
<td>22.9115</td>
<td>9.7870</td>
<td>5.2648</td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>2.3192</td>
<td>7.6620</td>
<td>5.3063</td>
<td>3.2773</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.5345</td>
<td>3.7060</td>
<td>3.1193</td>
<td>2.1760</td>
</tr>
<tr>
<td>n = 500, w = 10</td>
<td>0.05</td>
<td>374.0006</td>
<td>266.7122</td>
<td>283.8541</td>
<td>106.0184</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>11.3225</td>
<td>173.4994</td>
<td>48.2563</td>
<td>16.08</td>
</tr>
<tr>
<td></td>
<td>0.15</td>
<td>1.7788</td>
<td>10.1681</td>
<td>6.4022</td>
<td>3.3297</td>
</tr>
<tr>
<td></td>
<td>0.2</td>
<td>1.0569</td>
<td>1.9281</td>
<td>1.9134</td>
<td>1.4194</td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>1.0025</td>
<td>1.1281</td>
<td>1.1684</td>
<td>1.0741</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>1.0072</td>
<td>1.0178</td>
<td>1.0063</td>
</tr>
</tbody>
</table>

*n*: sample size, *w*: interval width, *δ* and *γ*: positive shifts in the process variance.

The reason for the low performance of the midpoint method in detecting the shift in the standard deviation is the fact that this method is over estimating it [57] as seen in Figure 28. In this figure, the interval widths on the horizontal axis are represented in terms of standard deviation units (*w* = 30, 24, 20, 15, 10, 5 corresponds to 0.2σ, 0.4σ, 0.6σ, 0.8σ, 0.9σ, 1.2σ respectively). The vertical axes show the expected bias in the sample mean and standard deviation estimations from the classic and midpoint methods which are the result of 5000 iterations with 1000 subgroup size.
The classic method estimates the population mean by averaging the actual subgroup observations (the data are not grouped at this point), while the midpoint method estimates the mean from the grouped data. The midpoint estimation of the mean is as accurate as classic method which is why the bias is pretty close to zero. The expected bias in the midpoint standard deviation increases as the number of intervals decreases.

4. Application to the VA Colonoscopy Services

Each method is applied to colonoscopy screening services, one of the most important specialty care services within the Veterans Health Administration (VA) network. A colonoscopy is the most costly and invasive screening option for Colorectal cancer (CRC): to prevent the spread of the disease and decrease the mortality rate, early detection and treatment of the disease is an important step [62]. The U.S. Preventive Services Task Force recommends using a high-sensitivity fecal occult blood testing (FOBT), sigmoidoscopy, or colonoscopy for men and women aged 50–75 [63]. A colonoscopy also serves as a follow-up test if anything unusual is found during one of the other screenings.

In 2008, a study based on data collected by the VA's Office of Quality and Performance stated that of the approximately 80% of Veterans received one of these screenings, nearly 72% underwent a colonoscopy within the last ten years [64]. However, in an analysis of FY 2002 data, the VHA Quality Enhancement
Research Initiative found that 54% of veterans with positive FOBT results failed to receive complete diagnostic evaluations within six months [65]. In order to reflect the time between FOBT testing and a follow-up colonoscopy, the previously discussed methods will be applied. The data set that is used in this application contains positive FOBT dates, colonoscopy dates, and the number of days elapsed until the colonoscopy. The sample time series data set is shown in Table 19 and the number of days to colonoscopy is plotted in Figure 29(a), which shows a right tailed dataset.

<table>
<thead>
<tr>
<th>Positive FOBT Date</th>
<th>Colonoscopy Performed</th>
<th>Colonoscopy Date</th>
<th>Days to Colonoscopy</th>
</tr>
</thead>
<tbody>
<tr>
<td>10/1/2010</td>
<td>1</td>
<td>11/30/2010 10:00</td>
<td>60</td>
</tr>
<tr>
<td>10/5/2010</td>
<td>1</td>
<td>12/3/2010 9:00</td>
<td>59</td>
</tr>
<tr>
<td>10/6/2010</td>
<td>1</td>
<td>11/30/2010 12:00</td>
<td>55</td>
</tr>
<tr>
<td>10/6/2010</td>
<td>1</td>
<td>12/2/2010 9:00</td>
<td>57</td>
</tr>
<tr>
<td>10/6/2010</td>
<td>1</td>
<td>12/2/2010 9:00</td>
<td>57</td>
</tr>
<tr>
<td>10/6/2010</td>
<td>1</td>
<td>12/3/2010 10:00</td>
<td>58</td>
</tr>
</tbody>
</table>

Figure 29. (a) Complete dataset of days to colonoscopy after positive FOBT, (b) Cumulative run chart for grouped data used by the VA.

The frequency distribution is presented partially in Table 20. The values in the first five columns represent frequencies of the colonoscopies in 30, 60, 90, 120, >120 days per month. The VA uses a cumulative run chart to observe this dataset which is in Figure 29(b). They prefer to use run charts for monitoring the processes due to its simplicity; however it is not always convenient. The run charts have
disadvantages including treating all movements with equal importance and leaving interpretation to the subjective position of whoever looked at the data. A control chart, essentially a graphical interpretation of a standard hypothesis test, is more useful and is comparably simple to a run chart. A quantile chart is a good fit for this type of dataset. The next columns contain estimations of the process mean and standard deviation for classic $\bar{x}$ and $s$ charts and the quantile chart. The resulting control charts are given in Figure 30.

Table 20. Frequencies of colonoscopy in 30, 60, 90, 120, >120 days, and sample mean and standard deviations estimated by classic and midpoint approach

<table>
<thead>
<tr>
<th>Month</th>
<th>30 days</th>
<th>60 days</th>
<th>90 days</th>
<th>120 days</th>
<th>&gt;120 days</th>
<th>xbar</th>
<th>stdev</th>
<th>midpoint mean</th>
<th>midpoint stdev</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oct-10</td>
<td>1</td>
<td>22</td>
<td>6</td>
<td>8</td>
<td>7</td>
<td>96.409</td>
<td>80.990</td>
<td>73.636</td>
<td>83.173</td>
</tr>
<tr>
<td>Nov-10</td>
<td>4</td>
<td>10</td>
<td>8</td>
<td>5</td>
<td>9</td>
<td>129.694</td>
<td>127.257</td>
<td>79.167</td>
<td>90.714</td>
</tr>
<tr>
<td>Dec-10</td>
<td>3</td>
<td>13</td>
<td>12</td>
<td>2</td>
<td>11</td>
<td>112.463</td>
<td>104.294</td>
<td>78.659</td>
<td>89.462</td>
</tr>
<tr>
<td>Jan-11</td>
<td>3</td>
<td>12</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>71.048</td>
<td>52.185</td>
<td>57.857</td>
<td>69.872</td>
</tr>
<tr>
<td>Jan-12</td>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feb-12</td>
<td>7</td>
<td>7</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>37.750</td>
<td>23.136</td>
<td>35.625</td>
<td>42.764</td>
</tr>
</tbody>
</table>

Figure 30. Monitoring (a) mean and (b) standard deviation with classic and midpoint approach

In the $\bar{x}$ chart in Figure 30(a), the estimated values of process mean and control limits seem very close to the actual values, although the process distribution is not distributed normally. In Figure 30(b), the estimated control limits look closer; however, the plotted standard deviation values are further from each other. These differences are due to the nonsymmetric shape of the dataset. In Figure 31(a), the chi-square
chart indicates an upper shift in the last periods. In order to understand which component(s) contributes to this shift, the chi-square table must be observed. According to the table given in Appendix 2, the observed frequency at 90 days is more than expected, while the observed frequency at 60 days is less than the expected value. The \( p \) chart in Figure 31(b) plots the proportions between days and colonoscopies that are less than 60, plus it demonstrates an in-control process.

![Figure 31. (a) Chi-square chart for multivariate monitoring, (b) \( p \) chart for proportion of days <60](image)

5. Discussion

In various fields, the observations are classified into intervals rather than exact values. Problems arise when the grouped dataset is used to make inferences about the parameter \( \theta \) because there is no complete information about the sample. According to the simulations, the quantile chart is an applicable alternative to the \( \bar{x} \) and \( s \) chart for presenting narrow interval widths and adequate sample sizes. Although the chi-square chart is a powerful and efficient way of multivariate monitoring, transforming a continuous data into discrete data decreases the three-sigma limit chi-square chart performance. However, it performs much better when probability limits are used. The chi-square chart with probabilistic limits is another alternative to monitor grouped data, especially for the small number of intervals and large sample size. The \( p \) chart is often used in the industry with a threshold. The simulations illustrate that unless the threshold is close to the actual value of process mean, the \( p \) chart does not perform well. In addition, all methods detect small shifts faster with large sample sizes and when actual data is distributed normally. As
a theoretical approach, the maximum likelihood estimation is complex due to the vast amount of
equations the fact that its application with Newton Raphson method does not always converge to a
solution. The minimum chi-square estimation method is easy to apply and yields close estimations to the
actual parameter values. Although these methods are applicable alternatives when the actual data is not
available, the grouped data has some limitations. The performances of the methods decrease when there is
not enough data. They are also affected by the sample size each method requires to achieve given error
rates, and the width of the intervals.
Conclusion

In many processes, regardless of how well-maintained, a certain amount of natural variability always exists. Control is necessary within this non-stationary world. The healthcare processes also have this systematic variation as well as special cause variation, which changes the process unnaturally. Like other statistical methods, Statistical Process Control (SPC) is used to tease out the inherent variability in a process. SPC methods have the advantage of integrating the power of statistical tests with analysis of graphs over time, and also detecting the changes and trends and with this data creating relatively simple formulas.

In order to improve healthcare performance, we must change our way of working. Control charts, the key to SPC, are used to visualize and analyze the performance of a process including biological processes such as anesthesia or organizational processes such as patient waiting time in a hospital. These tools help users to determine whether the performance of a process is stable and predictable or whether there is variation in the process, which creates an unstable and unpredictable process. Such variation can lead to an intervention in order to improve the process.

This study aims to illustrate the application of SPC methods in healthcare processes, particularly for adverse events and for the processes with grouped data, as well as examining the performances of certain methods to more quickly detect the process changes. This early detection would be beneficial within health care in many aspects such as cost reduction, lower mortality rates, more effective patient care, and so on. As an example, consider Central Line-Associated Bloodstream Infection (CLABSI) event. Central line is “an intravascular catheter that terminates at or close to the heart or in one of the great vessels which is used for infusion, withdrawal of blood, or hemodynamic monitoring” [66]. CLABSI is a primary laboratory confirmed bloodstream infection in a patient with a central line at the time of (or within 48-hours prior to) the onset of symptoms and the infection is not related to an infection from another site. An
estimated 41,000 CLABSI events occur in U.S. hospitals each year. These infections are usually serious infections typically causing longer hospital stay, increased cost and risk of mortality.

Wirtschafter et al uses SPC methods to detect significant change in the overall collaborative CLABSI rates among 13 collaborating regional neonatal intensive care units (NICU) in California [67]. The CLABSI rate per 1000 line days in the baseline period compared to the intervention and post intervention periods. They use a simple \( u \) chart to determine whether the collaborative was associated with systematic change in outcomes. They detected special cause variation, implying that the collaborative was associated with reduced infection rates, from 4.32 to 3.22 per 1000 line days (a 25% decrease) when comparing the baseline with the follow-up period. CLABSI rates for the 13 participating NICUs fell 25%, compared with the baseline, after the collaborative was started, which means approximately 75 fewer infections per year. CLABSIs are associated with significant morbidity and costs, thus this reduction is clinically significant. In 2011, the studies indicated that there were approximately 20,800 central line-acquired blood-stream infections in the United States per month (around 250,000 CLABSIs per year), with an estimated mortality rate of 1 in every 4 patients acquiring an infection [68]. Translating this frequent, preventable occurrence into a monetary figure shows that the estimated total cost of all CLABSIs in the United States ranges from $0.6 billion to $2.7 billion annually; assuming a conservative annual cost of $1.7 billion, the monthly cost of CLABSIs is approximately $142 million, where the cost of treatment for a CLABSI ranges from $6,000 to $29,000. However, using \( g \) charts to monitor CLABSI statistics, the capability to detect adverse events provides a tremendous opportunity for cost savings, as well as the prevention of thousands of mortalities. In the collaborative example above, they decreased the infections 25% which corresponds to 75 fewer infections among the collaborative per year. This means 300 infections annually and 25 infections monthly for the collaborative. Suppose that the CLABSI rate is shifted from the in-control \( p_0 = 0.1 \) to \( p = 0.11 \) with a 10% shift in the rate which means an average of 2.5 more infections occur in a month after the shift. According to the ANI values presented in Chapter 2, the \( g \) chart detects this shift almost two times faster than a \( p \) chart. Assuming the \( g \) chart can detect deviations
in the CLABSI time-between-infections two months earlier than a traditional \( p \) or \( u \) chart, every two months of faster detection can reduce CLABSI treatment costs by approximately $70,000, and associated mortality by one CLABSI-related death (since the estimated mortality rate is 1 in every 4 patients acquiring an infection).

This thesis shows that there are more powerful methods in order to detect this type of variation in the process more quickly. If we detect the changes faster, we can improve the process faster. While the first chapter illustrates the application of SPC methods to neonatology, the second chapter discusses the performances of the Shewhart and EWMA \( p \) and \( g \) charts for monitoring adverse healthcare event rates. The results suggest that EWMA \( g \) chart and probability-based limit \( g \) charts are more powerful for monitoring the CLABSI rates. The third chapter is about the Specialty Care Collaborative of the VA and the dashboard created for monitoring their baseline performance, which could lead to interventions for improvement. The final chapter explores the SPC methods to monitor grouped data sets, and suggests that the quantile chart is an applicable alternative to the \( \bar{x} \) chart when only the grouped data is available. Although statistical analyses are relatively easier when the actual data is available with high precision, the data might be classified into intervals and the actual values are not available. Another alternative for such data is the chi-square chart with probability limits. The given control charts perform better with a large sample size, narrower interval widths, and when the actual data comes from Normal distribution. As important as it is to use SPC methods in health care, it is equally important to use it correctly. SPC methods are highly desirable in rapidly detecting needed changes, for the cost of an undesirable process can be quite high.
References


1. The first derivative of Normal likelihood function for grouped data

\[
\frac{\partial \ln L}{\partial \mu} = \sum_{j=1}^{k-1} Q_j \left( \frac{e^{-\frac{1.5976(t_{j-1}-\mu)}{\sigma}} - \frac{0.07056(t_{j-1}-\mu)^3}{\sigma^3}}{1 + e^{-\frac{1.5976(t_{j-1}-\mu)}{\sigma}} - \frac{0.07056(t_{j-1}-\mu)^3}{\sigma^3}} \right)^2 - \frac{1}{1 + e^{-\frac{1.5976(t_{j-1}-\mu)}{\sigma}} - \frac{0.07056(t_{j-1}-\mu)^3}{\sigma^3}} \left( \frac{1.5976}{\sigma} + \frac{0.2117(t_{j-1} - \mu)^2}{\sigma^3} \right)
\]

\[
+ Q_k \frac{e^{-\frac{1.5976(t_{k-1}-\mu)}{\sigma}} - \frac{0.07056(t_{k-1}-\mu)^3}{\sigma^3}}{1 + e^{-\frac{1.5976(t_{k-1}-\mu)}{\sigma}} - \frac{0.07056(t_{k-1}-\mu)^3}{\sigma^3}}^2 \left( 1 - \frac{1}{1 + e^{-\frac{1.5976(t_{k-1}-\mu)}{\sigma}} - \frac{0.07056(t_{k-1}-\mu)^3}{\sigma^3}} \right)
\]

81
2. The second derivative of Normal likelihood function for grouped data

\[
\frac{\partial^2 \ln L}{\partial \mu^2} = Q_k \frac{e^{-1.5976(-\mu + t_{k-1})} \cdot 0.07056(-\mu + t_{k-1})^3}{(1 + e^{-1.5976(-\mu + t_{k-1})} \cdot 0.07056(-\mu + t_{k-1})^3)^2} (0. - \frac{0.42336(-\mu + t_{k-1})}{\sigma^3})
\]

\[
- \frac{e^{-3.1952(-\mu + t_{k-1})} \cdot 0.14112(-\mu + t_{k-1})^3}{(1 + e^{-1.5976(-\mu + t_{k-1})} \cdot 0.07056(-\mu + t_{k-1})^3)^4} (1 - \frac{1}{1 + e^{-1.5976(-\mu + t_{k-1})} \cdot 0.07056(-\mu + t_{k-1})^3})
\]

\[
+ \frac{2e^{-3.1952(-\mu + t_{k-1})} \cdot 0.14112(-\mu + t_{k-1})^3}{(1 + e^{-1.5976(-\mu + t_{k-1})} \cdot 0.07056(-\mu + t_{k-1})^3)^3} (1 - \frac{1}{1 + e^{-1.5976(-\mu + t_{k-1})} \cdot 0.07056(-\mu + t_{k-1})^3})
\]

\[
+ \frac{e^{-1.5976(-\mu + t_{k-1})} \cdot 0.07056(-\mu + t_{k-1})^3}{(1 + e^{-1.5976(-\mu + t_{k-1})} \cdot 0.07056(-\mu + t_{k-1})^3)^2} (1 - \frac{1}{1 + e^{-1.5976(-\mu + t_{k-1})} \cdot 0.07056(-\mu + t_{k-1})^3})
\]
\[- \sum_{j=1}^{k-1} Q_j \left( e^{-\frac{1.5976(-\mu + t_j)}{\sigma}} - \frac{0.07056(-\mu + t_j)^3}{\sigma^3} \left( \frac{1.5976}{\sigma} + \frac{0.21168(-\mu + t_j)^2}{\sigma^3} \right) \right) ^2 - e^{-\frac{1.5976(-\mu + t_{j-1})}{\sigma}} - \frac{0.07056(-\mu + t_{j-1})^3}{\sigma^3} \left( \frac{1.5976}{\sigma} + \frac{0.21168(-\mu + t_{j-1})^2}{\sigma^3} \right) ^2 \]

\[+ \left( e^{-\frac{1.5976(-\mu + t_j)}{\sigma}} - \frac{0.07056(-\mu + t_j)^3}{\sigma^3} \left( 0 - \frac{0.42336(-\mu + t_j)}{\sigma^3} \right) \right) ^2 \]

\[\left( 1 + e^{-\frac{1.5976(-\mu + t_j)}{\sigma}} - \frac{0.07056(-\mu + t_j)^3}{\sigma^3} \right) ^3 \]

\[+ e^{-\frac{3.1952(-\mu + t_{j-1})}{\sigma} - \frac{0.14112(-\mu + t_{j-1})^3}{\sigma^3} \left( \frac{1.5976}{\sigma} + \frac{0.21168(-\mu + t_{j-1})^2}{\sigma^3} \right) ^2 \]

\[- e^{-\frac{1.5976(-\mu + t_{j-1})}{\sigma}} - \frac{0.07056(-\mu + t_{j-1})^3}{\sigma^3} \left( 0 - \frac{0.42336(-\mu + t_{j-1})}{\sigma^3} \right) ^2 \]

\[\left( 1 + e^{-\frac{1.5976(-\mu + t_{j-1})}{\sigma}} - \frac{0.07056(-\mu + t_{j-1})^3}{\sigma^3} \right) ^3 \]

\[- e^{-\frac{3.1952(-\mu + t_{j-1})}{\sigma} - \frac{0.14112(-\mu + t_{j-1})^3}{\sigma^3} \left( \frac{1.5976}{\sigma} + \frac{0.21168(-\mu + t_{j-1})^2}{\sigma^3} \right) ^2} \]
\[ e^{-\frac{1.5976(-\mu + t_{j-1})}{\sigma} - \frac{0.07056(-\mu + t_{j-1})^3}{\sigma^3}} \left( \frac{1.5976}{\sigma} + \frac{0.21168(-\mu + t_{j-1})^2}{\sigma^3} \right)^2 \]

\[ \frac{(1 + e^{-\frac{1.5976(-\mu + t_{j})}{\sigma} - \frac{0.07056(-\mu + t_{j})^3}{\sigma^3}})^2}{1 + e^{-\frac{1.5976(-\mu + t_{j-1})}{\sigma} - \frac{0.07056(-\mu + t_{j-1})^3}{\sigma^3}}} \]
3. The second derivative of Weibull likelihood function for grouped data

\[
\frac{\partial^2 \ln L(\lambda, \beta, \gamma)}{\partial \lambda^2} = \sum_{j=1}^{k-1} Q_j \frac{1}{e^{\left(\frac{t_j - 1 - \gamma}{\lambda}\right)\beta} - e^{\left(\frac{t_j - \gamma}{\lambda}\right)\beta}} \left[ (t_{j-1} - \gamma)^2 (-1 + \beta) \beta e^{\left(\frac{t_{j-1}-\gamma}{\lambda}\right)\beta} \left(\frac{t_{j-1} - \gamma}{\lambda}\right)^{\beta - 2} + (t_{j-1} - \gamma)^2 \beta^2 e^{\left(\frac{t_{j-1}-\gamma}{\lambda}\right)\beta} \left(\frac{t_{j-1} - \gamma}{\lambda}\right)^{2\beta - 2} 
+ \frac{(-1 + \beta)\beta (t_j - \gamma)^2 e^{\left(\frac{t_j - \gamma}{\lambda}\right)\beta} \left(\frac{t_j - \gamma}{\lambda}\right)^{\beta - 2}}{\lambda^4} - \frac{\beta^2 (t_j - \gamma)^2 e^{\left(\frac{t_j - \gamma}{\lambda}\right)\beta} \left(\frac{t_j - \gamma}{\lambda}\right)^{2\beta - 2}}{\lambda^4} - \frac{2(t_{j-1} - \gamma)\beta e^{\left(\frac{t_{j-1}-\gamma}{\lambda}\right)\beta} \left(\frac{t_{j-1} - \gamma}{\lambda}\right)^{\beta - 1}}{\lambda^3} \right] 
+ \left[ Q_k \frac{(t_{k-1} - \gamma)^2 (-1 + \beta) \beta e^{\left(\frac{t_{k-1}-\gamma}{\lambda}\right)\beta} \left(\frac{t_{k-1} - \gamma}{\lambda}\right)^{\beta - 2}}{\left(1 - e^{\left(\frac{t_{k-1}-\gamma}{\lambda}\right)\beta}\right)\lambda^4} - \frac{(t_{k-1} - \gamma)^2 \beta^2 e^{\left(\frac{t_{k-1}-\gamma}{\lambda}\right)\beta} \left(\frac{t_{k-1} - \gamma}{\lambda}\right)^{2\beta - 2}}{\left(1 - e^{\left(\frac{t_{k-1}-\gamma}{\lambda}\right)\beta}\right)^2 \lambda^4} 
- \frac{(t_{k-1} - \gamma)^2 \beta^2 e^{\left(\frac{t_{k-1}-\gamma}{\lambda}\right)\beta} \left(\frac{t_{k-1} - \gamma}{\lambda}\right)^{2\beta - 2}}{\left(1 - e^{\left(\frac{t_{k-1}-\gamma}{\lambda}\right)\beta}\right)\lambda^4} + \frac{2(t_{k-1} - \gamma)\beta e^{\left(\frac{t_{k-1}-\gamma}{\lambda}\right)\beta} \left(\frac{t_{k-1} - \gamma}{\lambda}\right)^{\beta - 1}}{\left(1 - e^{\left(\frac{t_{k-1}-\gamma}{\lambda}\right)\beta}\right)\lambda^3} \right] \right]
\]
### 3. Chi-square table for days to colonoscopy

<table>
<thead>
<tr>
<th>Month</th>
<th>30 days</th>
<th>60 days</th>
<th>90 days</th>
<th>120 days</th>
<th>&gt;120 days</th>
<th>sum</th>
<th>Expected values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>30</td>
<td>60</td>
<td>90</td>
<td>120</td>
<td>&gt;120</td>
<td></td>
<td>days</td>
</tr>
<tr>
<td>Oct-10</td>
<td>1</td>
<td>22</td>
<td>6</td>
<td>8</td>
<td>7</td>
<td>44</td>
<td>3.718</td>
</tr>
<tr>
<td>Nov-10</td>
<td>4</td>
<td>10</td>
<td>8</td>
<td>5</td>
<td>9</td>
<td>36</td>
<td>3.042</td>
</tr>
<tr>
<td>Jan-11</td>
<td>3</td>
<td>12</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>21</td>
<td>1.775</td>
</tr>
<tr>
<td>Sep-11</td>
<td>5</td>
<td>26</td>
<td>8</td>
<td>1</td>
<td>2</td>
<td>42</td>
<td>3.549</td>
</tr>
<tr>
<td>Oct-11</td>
<td>0</td>
<td>21</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>30</td>
<td>2.535</td>
</tr>
<tr>
<td>Nov-11</td>
<td>2</td>
<td>9</td>
<td>19</td>
<td>4</td>
<td>2</td>
<td>36</td>
<td>3.042</td>
</tr>
<tr>
<td>Dec-11</td>
<td>6</td>
<td>5</td>
<td>10</td>
<td>4</td>
<td>0</td>
<td>25</td>
<td>2.113</td>
</tr>
<tr>
<td>Jan-12</td>
<td>5</td>
<td>16</td>
<td>12</td>
<td>0</td>
<td>0</td>
<td>33</td>
<td>2.789</td>
</tr>
<tr>
<td>Feb-12</td>
<td>7</td>
<td>7</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>16</td>
<td>1.352</td>
</tr>
<tr>
<td>sum</td>
<td>54</td>
<td>330</td>
<td>123</td>
<td>57</td>
<td>75</td>
<td>639</td>
<td></td>
</tr>
</tbody>
</table>