Evaluation of Scan Methods Used in the Monitoring of Public Health Surveillance Data

Shannon Elizabeth Fraker

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

Doctor of Philosophy
in
Statistics

William H. Woodall, Chair
Jeffrey B. Birch
Marion R. Reynolds, Jr.
G. Geoffrey Vining

October 31, 2007
Blacksburg, Virginia

Keywords: Anomaly detection, CUSUM charts, EWMA charts, Recurrence Interval, Scan Method, Time-to-Signal

Evaluation of Scan Methods Used in the Monitoring of Public Health Surveillance Data

Shannon Elizabeth Fraker

Abstract

With the recent increase in the threat of biological terrorism as well as the continual risk of other diseases, the research in public health surveillance and disease monitoring has grown tremendously. There is an abundance of data available in all sorts of forms. Hospitals, federal and local governments, and industries are all collecting data and developing new methods to be used in the detection of anomalies. Many of these methods are developed, applied to a real data set, and incorporated into software. This research, however, takes a different view of the evaluation of these methods.

We feel that there needs to be solid statistical evaluation of proposed methods no matter the intended area of application. Using proof-by-example does not seem reasonable as the sole evaluation criteria especially concerning methods that have the potential to have a great impact

in our lives. For this reason, this research focuses on determining the properties of some of the most common anomaly detection methods. A distinction is made between metrics used for retrospective historical monitoring and those used for prospective on-going monitoring with the focus on the latter situation. Metrics such as the recurrence interval and time-to-signal measures are therefore the most applicable. These metrics, in conjunction with control charts such as exponentially weighted moving average (EWMA) charts and cumulative sum (CUSUM) charts, are examined. Two new time-to-signal measures, the average time-between-signal events and the average signal event length, are introduced to better compare the recurrence interval with the time-to-signal properties of surveillance schemes. The relationship commonly thought to exist between the recurrence interval and the average time to signal is shown to not exist once autocorrelation is present in the statistics used for monitoring. This means that closer consideration needs to be paid to the selection of which of these metrics to report.

The properties of a commonly applied scan method are also studied carefully in the strictly temporal setting. The counts of incidences are assumed to occur independently over time and follow a Poisson distribution. Simulations are used to evaluate the method under changes in various parameters. In addition, there are two methods proposed in the literature for the calculation of the *p*-value, an adjustment based on the tests for previous time periods and the use of the recurrence interval with no adjustment for previous tests. The difference in these two methods is also considered. The quickness of the scan method in detecting an increase in the incidence rate as well as the number of false alarm events that occur and how long the method signals after the increase threat has passed are all of interest. These estimates from the scan method are compared to other attribute monitoring methods, mainly the Poisson CUSUM chart.

It is shown that the Poisson CUSUM chart is typically faster in the detection of the increased incidence rate.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would like to thank Olivia Grigg, Howard Burkom, Shabnam Mousavi, and Marianne Frisen for their helpful input in this work. Their advice and comments throughout the process have served to help shed light on the presented material. I would also like to thank Matt Shelburne for his advice with some of the programming aspects of this work.

Chapter 1

Introduction

Public health surveillance is the collection, analysis, and interpretation of data essential to disease and health-care related monitoring. With the demand to be aware of possible outbreaks of disease, or of regions of individuals that might be more predisposed to certain ailments, there is a need for statistical methods to monitor the abundant available data. These methods not only need to detect an outbreak and possibly where it is occurring, they also need to do so in a timely manner. This detection can, hopefully, be done while the increased threat is occurring and remedial action can be taken.

Industrial statistical process control (SPC) comprises a set of monitoring methods that have been widely used in industry for decades. The use of control charts such as the exponentially weighted moving average (EWMA) chart and the cumulative sum (CUSUM) chart

is common practice. The details of such charts are given in Roberts (1959) and Page (1954), respectively. The properties of these control charts have been studied in detail (see Reynolds and Stoumbos, 1999; and Steiner, 1999; for examples).

Public health officials also use control charts in disease monitoring (see Woodall 2006). Often, adjustments to the traditional control charts need to be made. This is due to the fact that disease monitoring involves study units, mainly humans, which cannot be assumed to be homogeneous as frequently as items studied in the industrial setting. Often adjustments for variability can be made based on several factors. For instance, surgical performance can be adjusted for patient risk factors such as age and pre-surgical health conditions. An overview of risk-adjusted charts was given by Grigg and Farewell (2004) while a more focused presentation of risk-adjusted CUSUM charts was given by Steiner et al. (2000). As discussed by Shmueli (2007), there can also be seasonal effects or other trends in the health-related data. This could lead one to use time series modeling of the data, such as was done by Reis and Mandl (2003).

Although the use of control charts is becoming more widespread in disease monitoring, use of scan statistics is a very popular method of outbreak detection. Woodall et al. (2007) provided an overview of many of these methods. When spatial information of outbreaks is not available, the use of scan statistics is closely related to the use of moving average control charts from industrial SPC. The issue that is of concern, however, is that the properties of these scan methods have not been thoroughly studied. Here, the evaluation of some of the most common disease monitoring methods will be done for the temporal case, often utilizing established comparison methods from industrial SPC. The goal is to determine how effective the currently

used monitoring methods are in detecting disease outbreaks and also how the methods behave during periods when there is no outbreak. In Chapter 2, several monitoring metrics will be discussed. In Chapter 3, some of the metrics from Chapter 2 will be compared in terms of their application to EWMA charts, CUSUM charts, and Markov dependent signaling processes. In Chapters 4 the performance of the temporal scan method proposed by Kulldorff (1997 and 2001) will be examined while in Chapter 5 the Kulldorff scan method will be compared to other monitoring methods. Finally, Chapter 6 will provide future research directions.

Chapter 2

Summary of Commonly Applied Surveillance Schemes

There are many methods that have been employed in the monitoring of disease rates. As with all statistical techniques, when a particular method can be used is dependent upon the underlying assumptions of the method. Unlike with many industrial monitoring methods, however, there has not been adequate evaluation of some disease monitoring schemes. Often times, a new technique is developed and then evaluated only in terms of a public health dataset that is complicated, with the timing and duration of any true outbreaks unknown. Because of this evaluation-by-example philosophy, how the method behaves, or why it works, is not always clear, especially in terms of the performance of the method during times of stability in the incidence rate. In this chapter, several metrics used in public health surveillance are discussed and their applicability to different surveillance schemes explained.

2.1:Retrospective Monitoring Metrics

Retrospective surveillance methods are designed for use when all data of interest have been collected. Often, this type of surveillance is used when detection methods are applied to an available dataset as if the surveillance was done at the time of the data collection. This type of surveillance is analogous to Phase I monitoring in industrial SPC.

The CDC (2001) and Buehler et al. (2004) recommended the use of sensitivity and predictive value positive (PVP) for evaluating the performance of a surveillance scheme. The metrics, however, are more commonly applied to the evaluation of the performance of a medical diagnostic procedure. The meaning of these metrics is quite clear when individual patients are undergoing diagnosis to determine whether or not a particular condition is present. Table 2.1 shows the classifications as shown in Table 3 of CDC (2001).

	Condition Present	Condition Absent	
Detected	A	B	$A+B$
No signal	C	D	$C+D$
	$A+C$	$B+D$	

Table 2.1: Classification used to calculate sensitivity and PVP

The sensitivity value is calculated as $A/(A+C)$. This metric is interpreted as the percentage of time a condition was correctly detected. The value of PVP is $A/(A+B)$ and is interpreted as the proportion of detections for which the condition under surveillance is present. For these metrics, higher values indicate a better diagnostic procedure, with perfect results corresponding to values of unity for both metrics.

Application of these metrics in surveillance, however, is less clear. Instead of individual patients being classified, the individual time periods of surveillance would be classified. At each time period there is either an outbreak detected or not and an outbreak is either present or not. One issue in the use of these metrics for surveillance data is that their values are invariant to the time order. Thus, two methods with strikingly different performance over a period of time can have the same values of these metrics.

As an illustrative example, consider a time period of length thirty. Suppose outbreaks occur over the time intervals 2-4, 9-12, and 23-30. Suppose Method 1 signals at time periods 7, 16, and 27-30; and Method 2, at time periods 2, 9, 16-17, and 23-24. These scenarios are illustrated in Figures 2.1 and 2.2, where values of unity for the indicator variable corresponding to the outbreak and to each method indicates the presence of an outbreak or a signal, respectively.

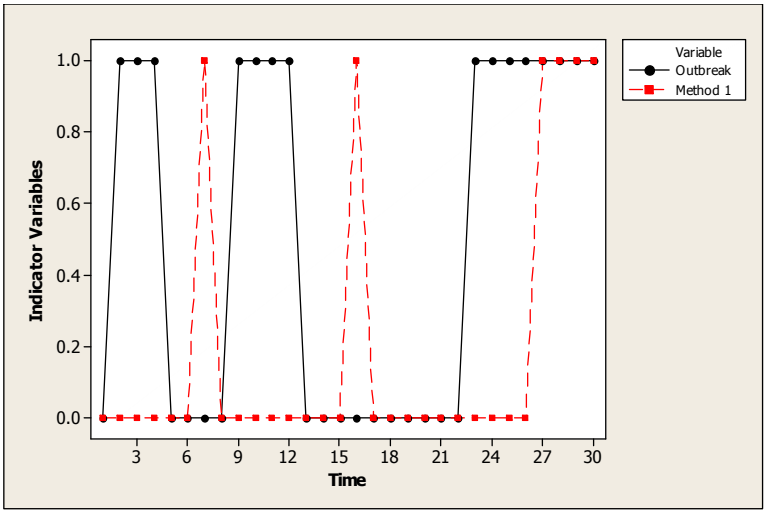


Figure 2.1: Plot of Outbreak with Signals from Method 1.

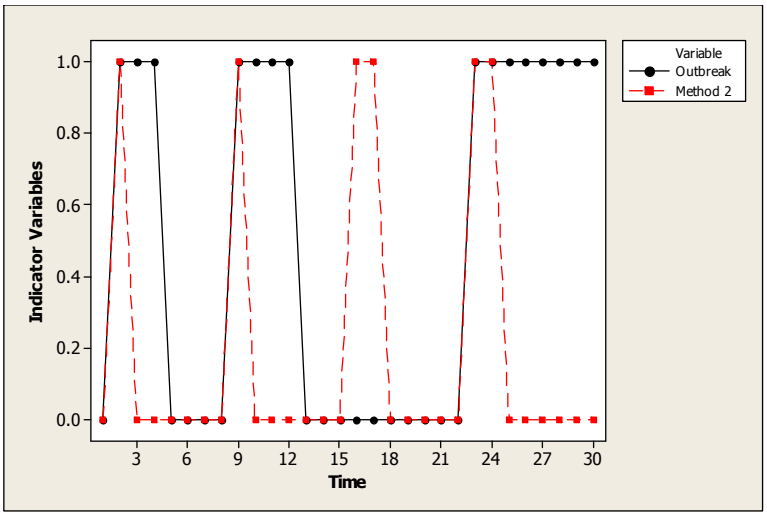


Figure 2.2: Plot of Outbreak with Signals from Method 2.

In each case we have the results summarized in the Table 2.2 with corresponding sensitivity, PVP, and PVN values of 4/15, 4/6, and 13/24, respectively. Method 2 has much better performance, however, because all of the outbreaks are detected immediately and the two false alarms occur successively. Method 1 fails to detect the first two outbreaks, has two separated false alarms, and detects the last outbreak more slowly than Method 1. This information is not evident in these metrics.

	Condition Present	Condition Absent	
Detected	4	2	6
No signal	11	13	24
	15	15	

Table 2.2: Classification used to calculate sensitivity and PVP in Example

A more reasonable definition of sensitivity was later given by Stoto et al. (2006). Sensitivity was defined as the probability that an outbreak will be detected when there is, in fact, an outbreak. Detection of an outbreak would, therefore, correspond to having an alarm occur at any point within the time period of the outbreak.

Stoto et al. (2006) also defined a false positive to be a time period for which a surveillance algorithm signals when there is no actual outbreak. This is typically referred to as a false alarm in the industrial process monitoring literature. Specificity is defined as 1 – P(false positive) or the probability that an alarm will not be raised at a time period when there is no outbreak. The use of the specificity metric is an attempt to characterize the in-control performance of a surveillance method, whereas the sensitivity metric is used to assess the power of a method to detect an outbreak. It is understood that the term in-control might be too strong of a term when it comes to health care monitoring. Using the term "in-control", however, simply refers to situations where there is no outbreak present.

2.2: Prospective Monitoring Metrics

Prospective surveillance is used when data are examined as they are being collected. Each time a new observation is recorded, the data must be reexamined to determine if an outbreak is believed to be occurring based on the new piece of information. This is analogous to Phase II monitoring in industrial SPC.

The use of metrics designed for retrospective surveillance in prospective monitoring is not recommended. As Frisén and Sonesson (2005) stated, "Evaluation by the significance level,

power, specificity and sensitivity which is useful for a fixed sample is not appropriate in a surveillance situation without a modification since they have no unique value unless the time period is fixed." Alternative metrics are, therefore, necessary for prospective monitoring. Since this research is based on prospective monitoring, the metrics discussed in this section will be the focus throughout the rest of the dissertation.

2.2.1: The Average Time-Between-Signal Events and the Average Signal Event Length

One issue present when comparing surveillance methods from industrial SPC and public health surveillance methods is the fact that in public health surveillance applications, data are collected over time with surveillance often continuing uninterrupted even after the occurrence of a signal. Because of this, the average time-to-signal (ATS), which is defined as the average number of time periods until the first signal, does not provide enough information. New time-to-signal measures, therefore, need to be introduced.

In the case of a one-sided control chart or surveillance method, we define a signal event as a sequence of consecutive signals. If, however, the process is being monitored with a two sided control chart or surveillance method where decreases as well as increases in the incidence rates are to be detected, then we define a signal event as a sequence of consecutive signals that either all indicate increases in the incidence rate or all indicate decreases in the incidence rate. This is similar to the idea of Zhu et al. (2005) who referred to a sequence of consecutive signals as a single "detection event". The in-control average time-between-signal events (ATBSE) is then defined as the average number of time periods between signal events when the process is in-control and measures how long on average it will take a control chart to signal again once a

signal has already occurred. The in-control average signal event length (ASEL) is defined as the average length of a signal event and measures how long on average a control chart will signal during consecutive time periods once a signal occurs. In the estimation of these two measures, any consecutive time periods that result in signals on opposite sides of the centerline of a two-sided control chart are considered to be associated with two signal events with no non-signal events in between.

2.2.2: The Recurrence Interval

The recurrence interval is defined to be the fixed number of time periods required for the expected number of false alarms in a monitoring process to be one. The recurrence interval is similar to the idea of a *p*-value in traditional hypothesis testing. If the recurrence interval is too extreme when compared to a cut-off value, then the result is considered too unlikely to have occurred under the assumed conditions.

The recurrence interval is commonly used in prospective spatiotemporal monitoring procedures for which count data are available at regular time intervals for a specified number of regions, *S*. In this situation, the recurrence interval was defined in Kleinman et al. (2004) and Kleinman (2005) to be the fixed number of time periods of surveillance required so that one would expect to observe exactly one *p*-value at least as small as the one observed in a particular region. Generalized linear models and generalized linear mixed models were used to predict the distribution of the count for each sub-region. Using the small area regression and testing (SMART) method, the *p*-value is calculated as the probability of observing a count at least as

high as the observed count given the predicted distribution for a particular region at each time period. The formula for the recurrence interval was then given as

$$RI = (p\text{-value} * S)^{-1}. \quad (2.1)$$

A signal is produced for any region if the recurrence interval exceeds a preset threshold. If desired, multiple thresholds can be used with the action taken dependent on which threshold the recurrence interval exceeds.

Kleinman (2005) also stated that a surveillance scheme can be developed for the sums of the counts in overlapping windows of up to r recent time periods. The assumption made is that the counts in the S regions at a specific time are independent. Therefore, the p-value for each of the S regions at each of the r time periods can be calculated. The recurrence interval is then given as

$$RI = (p - \text{value} * r * S)^{-1}. \quad (2.2)$$

An expression for the in-control ATS or the estimation of the distribution of the in-control ATS for a set of signal rules based on the sums of counts in overlapping windows is difficult to determine without using simulation because the statistics being used for monitoring will be dependent. The p-values will be correlated even if the counts within the regions are independent.

Use of the recurrence interval in public health surveillance is widespread. It is commonly applied to Center for Disease Control and Prevention's nationwide monitoring system *BioSense* (www.cdc.gov/biosense/). Kleinman (2005) stated that by using the recurrence interval, one can use any retrospective spatial clustering test repeatedly by remembering that the *p*-values of such a sequence of tests should not fall below a threshold α more often than once every $1/\alpha$ time periods. Because of this, he stated that the prospective monitoring case of Kulldorff (2001), discussed in detail in Chapter 4, was not necessary.

Scan methods with the use of recurrence intervals have also recently been used in syndromic surveillance. Syndromic surveillance is the analysis of data prior to a diagnosis where the data contain information on trends such as reported symptoms at hospitals or over-the-counter medicine sales. See, for example, Yih et al. (2004) and Daniel et al. (2005). Nordin et al. (2005) used simulation to assess the effectiveness of the model-based scan-based approach of Kleinman et al. (2005) to detect anthrax attacks. As mentioned in Woodall et al. (2007), however, there is commonly confusion between the recurrence interval and the in-control ATS. For example, Waller (2004) stated that the recurrence interval is "the expected time to observe as extreme a result under the model." Nordin et al. (2005) stated that the recurrence interval is "the expected number of days of surveillance needed for one such cluster of at least the observed magnitude to occur in the absence of any actual outbreaks." In addition, Yih et al. (2005) defined the recurrence interval as "the expected number of days of surveillance needed for a signal of at least the observed magnitude to occur, in the absence of any true outbreaks." These are all definitions of the in-control ATS, not the recurrence interval.

The use of the recurrence interval is limited by its lack of applicability in certain situations. In calculating the recurrence interval, the marginal probabilities of a signal at each time period, under the assumption of an in-control process, are required to be equal. If the surveillance method is in a signal state at time *i*, then let $E_i = 1$; otherwise let $E_i = 0$. The marginal probability of a signal at time *i* is then $P(E_i = 1)$. This probability differs from the run length probability at time *i*, defined as $P(E_i = 1 \cap E_j = 0 , j<i)$, and from the conditional probability of a signal at time *i* given no previous signals considered by Margavio et al. (1995), i.e. $P(E_i = 1 \mid E_j = 0, j<i)$.

If the marginal probabilities of a signal are not equal, then the number of time periods for which the expected number of false alarms is one is not fixed over time and the recurrence interval is not well-defined. Under the assumption of independent random variables, X_1, X_2, X_3, … observed over time, Shewhart-type charts with constant control limits do have constant marginal probabilities of signal. In this case, the recurrence interval and the in-control ATS are both equal to $P(\text{false alarm})^{-1}$. If, in addition, the observations are assumed to be normally distributed, then the EWMA chart with time-varying control limits will satisfy the restriction of a constant marginal probability of a signal if the control limits are set as a constant multiple of the standard deviation of the EWMA statistic from the centerline. In Chapter 3 the marginal probabilities of a signal of a one-sided CUSUM chart and of a Markov dependent signaling process are examined. The *p*-value approaches typically used with scan-based methods will be examined in Chapter 4 to see if they also meet the requirement of a constant marginal probability of a signal.

2.3 Scan Statistics

Scan statistics have been developed for the spatial, temporal, and spatiotemporal cases. The temporal case will be discussed in detail in Chapter 4. Brief descriptions of the spatial and spatiotemporal cases, however, are presented in this section.

2.3.1: The Spatial Case

In the spatial case, counts of incidences are observed overtime within a region of interest. These counts can be recorded at a specific location or they can be aggregated into counts within sub-regions. The scan statistic method proposed by Kulldorff (1997 and 2001) involves moving a circle over the region of interest. The location of the circle is allowed to vary so that at different points of monitoring different neighboring sub-regions are included in the circle. The circle is allowed to vary in radius with the maximum size of the circle being the one that contains no more than half of the observed counts. For count data that is aggregated by sub-region, the entire sub-region count is included in the circle if the centroid is included in the circle. This method allows for more flexibility in the size of a possible cluster than those of fixed radius methods. It should be noted, however, that when using the scan statistic proposed by Kulldorff, a circle of one radius will scan the entire region of interest and then a circle of another radius will scan the entire region of interest. This process will continue until all circles of possible radii have scanned the entire region. This method is, therefore, very computer intensive. The number of relevant circles, however, is limited. If a circle stays centered at the same place and the radius of the circle increases but no new observations are included inside the larger circle, then the surveillance statistic used with this method can not be larger than the surveillance statistic for the

smaller circle. Therefore, the approach is computationally tractable. In fact, the free software *SaTScan*™, is available for public use.

In the discussion of this method in Kulldorff 1997 and 2001, the model is assumed to be either the Poisson model or the Bernoulli model. The surveillance statistic used is the likelihood ratio statistic calculated conditionally on the total number of observed incidences, *N*. An unusually large value of the surveillance statistic would imply that the circle of sub-regions is an area of interest in terms of it being a possible cluster corresponding to significantly higher incidence rates.

2.3.2: Derivation of the Likelihood Ratio Surveillance Statistic

Kulldorff (1997) derived the likelihood ratio test statistic for his scan method of cluster detection for both the Bernoulli and the Poisson models in the spatial scenario. The likelihood under the null hypothesis that there is a constant background rate for the entire region of interest and under the alternative hypothesis that the rate in a specified sub-region is higher than in the remaining region of interest, i.e., that this sub-region represents an anomalous cluster, was found. In doing so, because the distributions are based on the counts for non-overlapping regions, the distribution of the number of incidences in the sub-region is independent from the distribution of the number of incidences in the entire region excluding the sub-region. Kulldorff (1997) used the likelihoods under the two hypotheses to derive the likelihood ratio statistic as an anomaly measure for candidate clusters.

Kulldorff's (1997) derivation of the likelihood ratio statistic for the Poisson model relies on "the density function of a specific point being observed at location x". The rate parameter for a specific point in continuous two-dimensional space, however, would be zero since under the Poisson process model the rate is proportional to the area of the region of interest. If the counts of incidences are not aggregated into the counts of incidences in a set of sub-regions, his derivation does not appear to hold. An alternate derivation, therefore, is given here.

As stated in Fraker et al. (2007), as far back as 1940 (Przyborowski and Wilenski, 1940) it has been known that if X_1 follows the Poisson distribution with a mean of m_1 and X_2 follows the Poisson distribution with a mean of m_2, then the conditional distribution of X_1 given $n = X_1 + X_2$ follows a binomial distribution with the probability of success equal to $m_1 / (m_1 + m_2)$, assuming X_1 and X_2 are independent.

To apply this result to the spatial case, we assume that the rate of incidence in the sub-region of interest is μ_2 while the rate of incidence in the remainder of the entire region under surveillance is μ_1. The entire region has area A and the sub-region of interest is assumed to have area $a < A$. We let X_a represent the observed number of incidences in the sub-region of interest, X_A represent the observed number of incidences in the entire region, and $X_{A-a} = X_A - X_a$ represent the observed number of incidences in the entire region excluding the sub-region. The random variables X_a and X_{A-a} are Poisson distributed with rates $a\mu_2$ and $(A-a)\mu_1$, respectively, where $A-a$ is the area of the entire region excluding the sub-region. The conditional probability of observing x incidences in the sub-region, given that the total number of incidences in the entire region is N is then given by

$$P(X_a = x \mid X_A = N) =$$

$$\binom{N}{x}\left[\frac{a\mu_2}{(A-a)\mu_1 + a\mu_2}\right]^x\left[1-\frac{a\mu_2}{(A-a)\mu_1 + a\mu_2}\right]^{N-x} \quad \text{for } x = 0, 1, 2, \ldots, N. \tag{2.3}$$

This expression is the binomial probability of observing exactly x incidences in the sub-region out of the N total number of incidences in the entire region of interest with a probability of an incidence occurring in the sub-region of

$$\left[\frac{a\mu_2}{(A-a)\mu_1 + a\mu_2}\right]. \tag{2.4}$$

If the rate of incidences in the sub-region is the same as the rate of incidences in the entire region excluding the sub-region, i.e., $\mu_1 = \mu_2$, then the expression in Equation (2.4) simplifies to a/A, and we have

$$P(X_a = x \mid X_A = N) = \binom{N}{x}\left(\frac{a}{A}\right)^x\left(1-\frac{a}{A}\right)^{N-x} \quad \text{for } x = 0, 1, 2, \ldots, N. \tag{2.5}$$

To calculate the likelihood ratio surveillance statistic, the likelihood under the alternative hypothesis of unequal incidence rates needs to be maximized with respect to the probability of being in the sub-region. This involves maximizing Equation (2.3) with respect to the sub-region incidence probability given in the expression in Equation (2.4). If we set the expression in

Equation (2.4) equal to p, then maximizing Equation (2.3) yields the maximum likelihood estimator in the simple binomial case, which is well-known to be

$$\hat{p} = \frac{x}{N}. \tag{2.6}$$

By substituting Equation (2.6) into Equation (2.3) for the probability of being in the sub-region, we obtain

$$\binom{N}{x}\left(\frac{x}{N}\right)^{x}\left(1-\frac{x}{N}\right)^{N-x} \tag{2.7}$$

We define $E_s = \left({}^{a}\!/\!_{A} \right) N$ to be the expected number of incidences in the sub-region of interest when there is no increase in the rate of incidences in this sub-region. The likelihood ratio surveillance statistic is then the ratio of the expression in Equation (2.7) to that in Equation (2.5), which results in

$$LR = \frac{L(x)}{L_o} = \frac{\binom{N}{x}\left(\frac{x}{N}\right)^{x}\left(1-\frac{x}{N}\right)^{N-x}}{\binom{N}{x}\left(\frac{a}{A}\right)^{x}\left(1-\frac{a}{A}\right)^{N-x}} = \left(\frac{x}{\left(\frac{a}{A}\right)N}\right)^{x}\left(\frac{N-x}{\left(\frac{A-a}{A}\right)N}\right)^{N-x} = \left(\frac{x}{E_s}\right)^{x}\left(\frac{N-x}{N-E_s}\right)^{N-x} \tag{2.8}$$

if $x > E_s$ and $L(x)/L_0 = 1$ otherwise for $x = 0, 1, 2, \ldots, N$. This surveillance statistic can also be used for situations when it is desired to detect a decrease in the rate. In this case, Equation (2.8)

is defined for $x < E_s$ and is equal to 1 otherwise. In either situation, Equation (2.8) is the same as Equation (2) in Kulldorff (2001). By taking the approach of starting with a conditional probability function, however, a simpler and more general derivation of the likelihood ratio surveillance statistic for the commonly used Poisson model is shown.

2.3.3: The Spatiotemporal Case

Kulldorff (2001) extends the results of Kulldorff (1997) to the prospective spatiotemporal case. Not only is the location of a cluster of interest, but the time of the cluster is also of interest. Instead of a circle of varying radius scanning the region of interest as in the spatial case, a cylinder scans the region of interest in the spatiotemporal case. Here, the base of the cylinder is accounting for the location and operates the same as the circle in the spatial case. The height of the cylinder accounts for time. The cylinder is allowed to vary in radius and in height under the restriction that no more than half of the observed incidences are in any given cylinder. The start time of the cylinder varies in both the retrospective and prospective cases. In the prospective monitoring case, however, the only cylinders of interest are those that end at the current time period. Kulldorff (2001) recommended using simulation to evaluate the significance of an observed value. Again the method is conditional on N, the total number of incidences at the current time. The details of this simulation are explained in terms of the temporal case in Chapter 4.

2.3.4: Discussion

Most of the performance comparisons of scan methods have been done for the retrospective case (see Kulldorff, Tango, and Park 2003 and Song and Kulldorff 2003). Even

though Kulldorff's method is widely used prospectively there have been only very limited statistical performance evaluations. Kulldorff (2001) introduced his scan method for prospective monitoring and then applied it to a complicated dataset and assumed that the clusters signaled as significant are just that, anomalies. There is no discussion of the properties of this method. This is not unusual in the public health surveillance literature, however, even though it would be valuable to better understand the statistical performance of the methods used.

In their approach to evaluating the spatiotemporal scan statistic, Kulldorff et al. (2004) simulated datasets based on the geography and population of New York City. Outbreaks of varying size and intensity were created. These datasets were posted at www.satscan.org/datasets with the aim that they could be used to evaluate new methods as they are proposed.

It is important to understand clearly the way in which these datasets were created. A number of different datasets were generated with each corresponding to either 31, 32, or 33 days of counts of incidences for each zip code region in the city. Separate simulations were performed for each number of days. In each case any outbreak occurred after the thirtieth day. The total number of observed cases was fixed to be 100 times the total number of days considered, with the random assignment of the cases to the regions and days made proportional to the population of the zip code region. The outbreak-affected areas had a higher probability based on the relative risk associated with the outbreak. The power of the scan method to detect each outbreak for a particular number of days considered was then determined based on repeated simulations. The power is thus defined as the probability of signaling an outbreak at time T given an outbreak at a specified earlier time period. This procedure allows one to estimate the marginal probabilities of

a signal at given times after an outbreak, but these probabilities cannot be used to determine the probabilities that the first signal occurs at given times. These latter probabilities are needed to study how quickly an outbreak can be detected.

There are several issues associated with this approach that limit its usefulness. Consider their simulations where 31 days of incidence counts were obtained. In this case there were always 3100 total cases. If there is an increased risk of cases occurring on day 31, then fewer cases will tend to occur in the first thirty-day period. The greater the level and scope of any outbreak, the lower the number of cases in the pre-outbreak period will tend to be. It would seem more realistic for this application to assume that the rate of cases occurring prior to any outbreak remains constant until the outbreak, regardless of the size of the outbreak.

More importantly, however, simulating data separately and independently for each day limits the information that can be gained from the approach. In applications one must use the data at day 31, for example, in the analysis required for monitoring at day 32. There is dependence over time that is ignored with the use of separate simulations for each number of days. The probabilities of detecting an outbreak within specified numbers of days are of most interest. This information simply cannot be obtained with independent simulations for each day to estimate the marginal probabilities of a signal.

Kulldorff (2001) did note a distinction between the control charting methods and the scan methods for the spatiotemporal case. He stated that the control chart type methods "… are geared towards detecting the sudden appearance of clustering as a global phenomenon throughout the

geographical area under study, as opposed to the currently proposed method, which is used to detect the specific geographical location of emerging clusters and to evaluate their significance. The two types of methods complement each other, being suitable for different types of inquiries." No methods for prospective cluster detection with spatiotemporal data have ever been proposed in the industrial SPC literature. Industrial SPC methods, however, are readily available for temporal data and therefore will be used in the evaluation of the scan methods proposed by Kulldorff for the temporal case.

Chapter 3

Comparison Between the Recurrence Interval and Time-to-Signal Properties of Surveillance Schemes

In this chapter, EWMA charts, CUSUM charts, and Markov dependent signaling processes that are allowed to continue uninterrupted even after a signal has occurred will be studied. This allows for the time-to-signal properties of these surveillance schemes to be compared with the recurrence interval for the temporal monitoring case, where data are available at regular time intervals.

3.1: The Recurrence Interval and the EWMA Chart

This section will focus on the non-resetting EWMA chart. Since it is known that the marginal probabilities of a signal can be constant for this chart, the recurrence interval will be calculated and compared to the time-to-signal measures.

3.1.1: Properties of the Non-resetting EWMA Chart

All previous studies of the properties of an EWMA chart have been under the condition that the chart is reset to the initial conditions after a signal has occurred, as is typically the case in industrial SPC. To compare the properties of the EWMA chart to the recurrence interval, however, the control chart needs to run continuously without interruption since the recurrence interval is defined for any point in time of the monitoring process as long as the process remains in-control. A simulation was, therefore, conducted to examine the properties of an EWMA chart that is not reset to its initial conditions after a signal has occurred. It is assumed throughout the simulation that observations, X_1, X_2, X_3,…, collected over time are the result of an in-control process. These observations are generated independently and identically from a normal distribution. Without loss of generality, it can be assumed that the observations are from the standard normal distribution. Once an observation is generated, the EWMA chart statistic is calculated using

$$M_i = \lambda X_i + (1-\lambda) M_{i-1} \tag{3.1}$$

where i = 1, 2, 3, …; λ is the smoothing constant, and M_0 is set to be the in-control mean, in this case zero. Commonly applied EWMA smoothing constant values of 0.05, 0.10, 0.20, 0.30, and 0.40 are used. The control limits applied are the three-sigma time-varying limits given by

$$\pm 3\left(\frac{\lambda}{2-\lambda}\left[1-(1-\lambda)^{2i}\right]\right)^{1/2} \tag{3.2}$$

Thus, the EWMA chart signals if the value of M_i falls beyond three standard deviations of the in-control mean. Note that for the EWMA chart, if the approximate control limits given by

$$\pm 3\left[\frac{\lambda}{(2-\lambda)}\right]^{1/2} \tag{3.3}$$

had been used instead of the time varying limits, then the marginal probabilities of a signal at each time point would not be constant and the recurrence interval would not be well defined.

The number of observations between signal events, as well as the length of each signal event, were calculated in a simulation study. The in-control ATS was estimated using the average number of time periods before the first signal event. The in-control ATBSE was estimated by the average of the numbers of observation between signal events and the ASEL was estimated by the average length of signal events. In the simulation, the relevant statistics were found for a control chart containing 10,000 time periods of observations. There were 2,500 such charts examined for each value of λ. It should be noted that since each chart was truncated at 10,000 observations, the ATBSE was calculated using all sequences of observations between signal events prior to the last signal event.

Since the two-sided EWMA chart is being used, the definition of a signal event is based on consecutive time periods resulting in a signal on the same side of the center line. It should be noted, however, that in the simulation of the EWMA chart, two consecutive signals never occurred on the opposite side of the center line. This is because the probability of having consecutive time periods with signals on opposite sides of the chart is very low. An upper bound

on the conditional probability of a false alarm occurring on one side of an EWMA control chart given a false alarm occurred on the opposite side of the center line during the previous time period can be calculated as follows:

$$P(M_i < -3*se_i \mid M_{i-1} = 3*se_{i-1}) =$$

$$P\left(\lambda X_i + (1-\lambda)M_{i-1} < -3\left[\frac{\lambda}{2-\lambda}[1-(1-\lambda)^{2i}]\right]^{1/2} \mid M_{i-1} = 3\left[\frac{\lambda}{2-\lambda}[1-(1-\lambda)^{2(i-1)}]\right]^{1/2}\right) =$$

$$P\left(\lambda X_i + 3(1-\lambda)\left[\frac{\lambda}{2-\lambda}[1-(1-\lambda)^{2(i-1)}]\right]^{1/2} < -3\left[\frac{\lambda}{2-\lambda}[1-(1-\lambda)^{2i}]\right]^{1/2}\right) =$$

$$P\left(X_i < \frac{-3\left[\frac{\lambda}{2-\lambda}[1-(1-\lambda)^{2i}]\right]^{1/2} - 3(1-\lambda)\left[\frac{\lambda}{2-\lambda}[1-(1-\lambda)^{2(i-1)}]\right]^{1/2}}{\lambda}\right) =$$

$$P\left(X_i < -3\left(\left[\frac{1-(1-\lambda)^{2i}}{\lambda(2-\lambda)}\right]^{1/2} + \left[\frac{(1-\lambda)^2-(1-\lambda)^{2i}}{\lambda(2-\lambda)}\right]^{1/2}\right)\right) \qquad (3.4)$$

where M_i is the EWMA chart statistic given in Equation (3.1) and se_i is the standard deviation of the EWMA statistic M_i. In all of the simulations, X_i is distributed as a standard normal random variable. Therefore, this expression can be evaluated for any value of the smoothing constant and for i = 2, 3, 4, … In particular, an interval on the conditional probability can be found by evaluating the expression at $i = 2$ *and* $i = \infty$. The value of

$$-3\left(\left[\frac{1-(1-\lambda)^{2i}}{\lambda(2-\lambda)}\right]^{1/2}+\left[\frac{(1-\lambda)^{2}-(1-\lambda)^{2i}}{\lambda(2-\lambda)}\right]^{1/2}\right) \tag{3.5}$$

is given in Table 3.1 for each of these choices of *i*. The probability of a standard normal random variable being less than any of these values is zero to at least seven decimal places. Therefore, it is highly unlikely that two false alarms will occur consecutively on opposite sides of the EWMA centerline when the three standard error time varying limits are used.

λ	$i = 2$	$i = \infty$
0.05	-6.9879	-18.7350
0.10	-6.7360	-13.0767
0.20	-6.2419	-9.0000
0.30	-5.7620	-7.1414
0.40	-5.2986	-6.0000

Table 3.1: Z-values for consecutive false alarms in an EWMA chart.

As stated previously, for an EWMA chart, the marginal probabilities of a false alarm are constant for all time periods as long as the time-varying control limits are set as a constant multiple of the standard deviation of the EWMA statistic. For instance, if this constant is three, as in our simulations, then $P(E_i = 1) = 0.0027$ for all time periods and the recurrence interval is $(0.0027)^{-1} = 370.4$ for any value of the smoothing constant. While the marginal probabilities of a signal and the recurrence interval are constant over λ, the time-to-signal measures are not. Figure 3.1 shows the estimated in-control ASEL, Figure 3.2 shows the estimated in-control ATBSE, and

Figure 3.3 shows the estimated in-control ATS. Also shown on the graphs are the theoretical values for an EWMA chart with $\lambda = 1.00$ (dashed line), which is equivalent to a Shewhart chart. For the in-control ASEL the largest standard error was 0.016, for the in-control ATBSE the largest standard error was 6.7, and for the in-control ATS the largest standard error was 26.0. For all cases, the largest standard error occurs at $\lambda = 0.05$. As the value of the smoothing constant decreases all three time-to-signal measures increase due to increased positive autocorrelation in the values of the EWMA statistics.

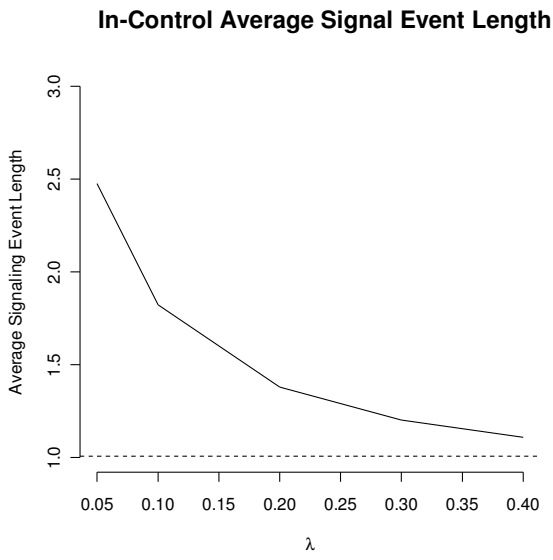


Figure 3.1: The in-control average signal event length. The solid line represents the simulated value and the dashed line represents the value of the average signal length when $\lambda = 1.00$ which is $\frac{1}{0.9973}$.

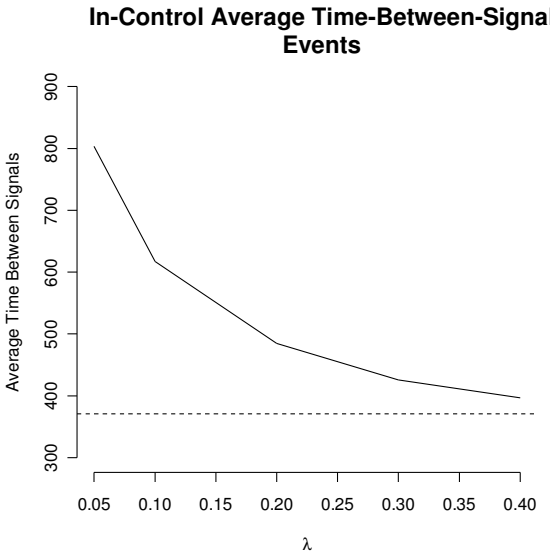


Figure 3.2: The in-control average time-between-signal events. The solid line represents the simulated value and the dashed line represents the value of the in-control ATBSE when $\lambda = 1.00$, which is the recurrence interval of 370.4.

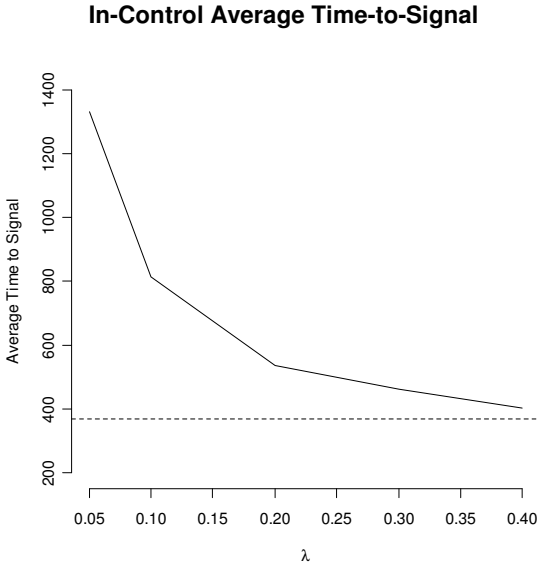


Figure 3.3: The in-control average time-to-signal. The solid line represents the simulated value and the dashed line represents the value of the in-control ATS when $\lambda = 1.00$, which is the recurrence interval of 370.4.

Also calculated in this simulation was the probability distribution function of the time-between-signal events for length ten or less. Figure 3.4 shows this distribution. As λ increases, the probability distribution function values in this range decrease and approach a limiting distribution. This limiting distribution, which occurs when λ = 1.0, is the geometric distribution given by

$$P(TBSE = k) = (0.0027)(0.9973^{k}), \tag{3.6}$$

where k = 0, 1, 2,…. As λ decreases, the EWMA chart is more likely to have a small value for the time-between-signal events even though the ATBSE increases as λ decreases, due to the fact that the distribution of the time-between-signal events is highly skewed. The relationship between the time-between-signal events and λ is explained by the fact that as λ decreases the degree of positive autocorrelation between consecutive EWMA statistics increases.

Table 3.2 gives the values of the estimated in-control ATS, estimated in-control ATBSE values, estimated in-control ASEL, the recurrence interval and, where available, Steiner's (1999) analytically determined in-control ATS values for all λ values under consideration. As shown in Table 3.2, the in-control ATS value is not equivalent to the in-control ATBSE value. For example, in the simulation results for λ = 0.05 the estimated in-control ATS is 1332.1 with a standard error of 26.0 and the estimated in-control ATBSE is 746.6 with a standard error of 6.7. Steiner (1999) used a Markov chain based approach and found the in-control ATS for this chart to be 1353.0. As stated before, the recurrence interval for the EWMA chart with three-sigma

time-varying limits is 370.4. Steiner (1999) also determined the in-control ATS for $\lambda = 0.25$ and $\lambda = 0.50$ to be 500.0 and 382.0, respectively.

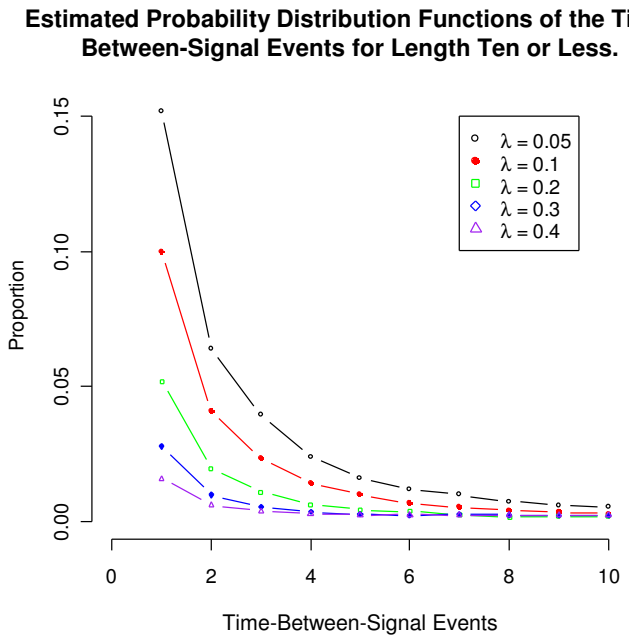


Figure 3.4: The probability distribution functions of the time-between-signal events for length ten or less for an EWMA chart.

The recurrence interval is constant, but the time-to-signal properties of the EWMA chart vary widely depending on the value of λ. For all values of λ the in-control ATBSE values are less than the in-control ATS values. The closeness of these performance measures at the larger λ values is due to the fact that both the in-control ATS and the in-control ATBSE are approaching the recurrence interval as the value of the smoothing constant increases. This relationship between the in-control ATS and in-control ATBSE implies that once a false alarm occurs in an EWMA chart, the average time to another signal would be less than it would be if the EWMA statistic were at the centerline, due to inertial effects. It can also be seen that as the amount of

positive autocorrelation in the EWMA statistic increases, the ASEL increases. These results are very reasonable intuitively.

λ	*Estimated In-Control ATS*	*Estimated In-Control ATBSE*	*Estimated In-Control ASEL*	*Steiner's ATS*	*Recurrence Interval*
0.05	1332.1 (26.0)	746.6 (6.7)	2.48 (0.016)	1353	370.4
0.10	814.9 (16.3)	603.0 (4.0)	1.82 (0.008)	828	370.4
0.20	537.5 (10.7)	482.6 (2.4)	1.38 (0.004)	-----	370.4
0.30	463.8 (9.4)	423.1 (1.9)	1.20 (0.002)	-----	370.4
0.40	403.1 (8.0)	396.0 (1.7)	1.11 (0.001)	-----	370.4

Table 3.2: Performance metric summary for the EWMA chart. Estimated in-control ATS, estimated in-control ATBSE, and estimated in-control ASEL values from the simulation results, in-control ATS values from Steiner (1999), and the recurrence interval values for the EWMA chart with three-sigma time-varying limits. Standard error values are given in parentheses.

3.1.2: The Recurrence Interval Applied to the EWMA chart

Since the recurrence interval is defined to be the fixed number of time periods for which the expected number of false alarms in a monitoring procedure is one, a second simulation was performed to find the distribution of the number of signals that occur in an EWMA chart for a length of time equal to the recurrence interval. Two different methods to do this are illustrated. In the first method, 10,000 charts were simulated under the same assumptions as the previous simulation. The number of signals that appear in the first 370 consecutive time periods was calculated for each of the 10,000 charts.

Table 3.3a shows the estimated proportions of times that no false alarms occurred, one false alarm occurred, and more than one false alarm occurred in the first 370 time periods for the

λ values under consideration, as well as the exact probabilities for λ = 1.0, based on the binomial distribution with n = 370 and p = 0.0027. As λ decreases, the probability of no false alarms in 370 time periods increases while the probability of one false alarm and the probability of more than one false alarm decrease. The proportion of no false alarms, however, is always the largest value except when λ = 1.0, i.e. when there is no autocorrelation in the monitoring statistic. The probability of exactly one false alarm is low and the value decreases as there is more autocorrelation in the control chart statistic. Therefore, using an expected value of one false alarm in 370 time periods, i.e. the approximate recurrence interval, does not appear to be particularly representative of what is occurring in the surveillance process. As a result, the recurrence interval does not adequately summarize the performance of the surveillance scheme. In this case, the average is not the most typical value.

In a second approach, again 10,000 charts were simulated under the same conditions as the previous simulation. Instead of considering the number of signals in the first 370 time periods for all of the 10,000 charts, the charts are divided into two groups. The first group contains 6,300 charts and the second group contains 3,700 charts. The number of signals that appear in the first 370 time periods for the first group and the number of signals that appear in the first 371 time periods for the second group were calculated. This second approach is a way to represent an approximate recurrence interval of 370.37.

Table 3.3b shows the estimated proportions of times that no false alarms occurred, one false alarm occurred, and more than one false alarm occurred in either the first 370 or 371 time

periods for the λ values under consideration, as well as the estimated probabilities for λ = 1.0. The conclusions are the same as those from Table 3.3a.

λ	*Proportion with no false alarms*	*Proportion with one false alarm*	*Proportion with more than one false alarm*
0.05	0.7547 (0.0043)	0.0796 (0.0027)	0.1657 (0.0037)
0.10	0.6369 (0.0048)	0.1466 (0.0035)	0.2165 (0.0041)
0.20	0.5142 (0.0050)	0.2391 (0.0043)	0.2467 (0.0043)
0.30	0.4487 (0.0050)	0.2933 (0.0046)	0.2580 (0.0044)
0.40	0.4083 (0.0049)	0.3282 (0.0047)	0.2635 (0.0044)
1.0	0.3678	0.3684	0.2639

Table 3.3a: Estimated false alarm probabilities for an EWMA chart with approximate RI = 370. The exact probabilities when λ = 1.00 are also included. Standard error values are given in parentheses.

λ	*Proportion with no false alarms*	*Proportion with one false alarm*	*Proportion with more than one false alarm*
0.05	0.7529 (0.0043)	0.0797 (0.0027)	0.1674 (0.0037)
0.10	0.6347 (0.0048)	0.1478 (0.0035)	0.2175 (0.0041)
0.20	0.5150 (0.0050)	0.2370 (0.0043)	0.2480 (0.0043)
0.30	0.4511 (0.0050)	0.2908 (0.0045)	0.2581 (0.0044)
0.40	0.4143 (0.0049)	0.3171 (0.0047)	0.2686 (0.0044)
1.0	0.3658 (0.0048)	0.3681 (0.0048)	0.2661 (0.0044)

Table 3.3b: Estimated false alarm probabilities for an EWMA chart with approximate RI = 370.37. Standard error values are given in parentheses.

The reason that both of these approaches were employed, is due to the fact that the recurrence interval is a non-integer value. One, however, can not sample fractions of time periods or fractions of observations. How to handle a recurrence interval that is not an integer value is, therefore, unclear. The two approaches present two different ways to handle a non-integer valued recurrence interval. In the first approach, the recurrence interval is rounded to the nearest integer value and the simulation is then applied in the same way for all charts. In the second approach, the simulation is applied differently for a different number of charts which allows the simulation to represent a recurrence interval that is a non-integer value. The number of decimal places chosen in this case is two which gave a recurrence interval approximation of 370.37. A recurrence interval approximation of 370.4, the estimate stated throughout the rest of this chapter, could also have been applied. In this case, instead of dividing the 10,000 simulated charts into groups of 6,300 and 3,700, the groups would be of size 6,000 and 4,000. The results of this simulation are also not significantly different from the other two. Therefore, none of these approximations result in significantly different probabilities.

It should also be noted that the probability of a signal being used throughout the chapter, 0.0027, is also an approximation. This approximation is also not expected to change the results considerably. Therefore, it seems that no matter how one chooses to approximate the recurrence interval when it is not an integer value, it does not summarize the time-to-signal performance of the surveillance method due to the positive autocorrelation in the EWMA values over time.

3.2: The Recurrence Interval and the CUSUM Chart

As with the use of EWMA charts in industrial SPC, CUSUM charts are usually reset to the initial conditions after a process adjustment is done following a signal. Therefore, a simulation similar to the EWMA control chart simulations was run to examine the properties of a one-sided CUSUM chart that is allowed to run uninterrupted without being reset to its initial conditions after a signal occurs. It is again assumed throughout the simulation that the observations are from an in-control process. Observations, X_1, X_2, X_3, …, are collected over time from a standard normal distribution. Once a point is generated, the one-sided CUSUM chart statistic is calculated using

$$S_i = \max(0, S_{i-1} + X_i - k) \tag{3.7}$$

where $i = 1, 2, 3, \ldots$ and $S_0 = 0$. A signal is given as soon as $S_i > h$.

Values of h of 4.0 and 5.0 and values of k of 0.5 and 1.0 were chosen, all of which are commonly used settings for the one-sided CUSUM chart. These choices of k correspond to wanting to quickly detect a shift of one or two standard deviations in the mean, respectively. Unlike with the EWMA chart, it was not known to what extent the marginal probabilities of a signal were constant. Since the recurrence interval can not be applied unless they are, the proportion of charts that result in a signal at each time period was calculated. These proportions are the estimated marginal probabilities of signal at each time point. For $k = 0.5$, there were 100,000 charts simulated, each for 100 time periods. Since the probabilities of a signal at each

time period are much smaller when $k = 1.0$, there were 1,000,000 charts simulated, each for 100 time periods in order to reduce the standard error to a reasonable value.

Figures 3.5, 3.6, 3.7, and 3.8 show the proportion of signals at each time point for all combinations of h and k under study. In each figure, the average proportion for values beyond time period 20 are shown (dashed line) as well as the three standard error bounds (solid lines), based on the average proportion after time 20. A time period of 20 was selected for estimating the average marginal probability of a signal since it appears that, overall, the proportions level off after that time period. The only exception to this is when $h = 5.0$ and $k = 1.0$; in this situation, the relative frequency for time period 71 falls just outside of the three standard error limits. This is likely due to random noise. For both values of h, when $k = 1.0$ the probabilities level off sooner than when $k = 0.5$. Prior to reaching the limit, there is a steady increase in the estimated marginal probabilities of signal. This implies that a one-sided CUSUM chart is less likely to signal at the beginning of the monitoring phase. The marginal probabilities of a signal are, therefore, not constant. Consequently, the recurrence interval can not be applied to a one-sided CUSUM chart unless one excludes the earlier time periods and considers steady-state behavior, similar to the approach of Grigg and Spiegelhalter (2006). The larger the value of the shift that is desired to be detected, the shorter the delay in reaching reach steady-state. To make it possible to calculate the recurrence interval for a one-sided CUSUM chart without considering the steady-state condition, the limits would need to be modified so that they are lower for earlier time periods. This would result in time-varying limits similar to those for the EWMA chart.

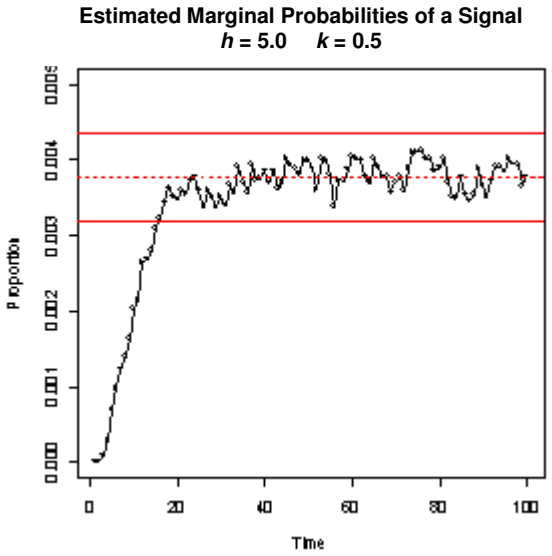


Figure 3.5: The estimated marginal probabilities of a signal for CUSUM parameter values of h = 5.0 and k = 0.5. The dotted solid line represents the simulated values; the dashed line, the average of the values after time 20; and the solid line, the three standard error bounds around this proportion.

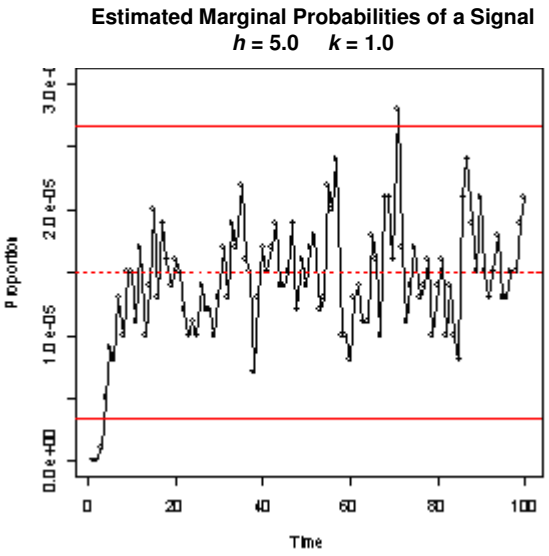


Figure 3.6: The estimated marginal probabilities of a signal for CUSUM parameter values of h = 5.0 and k = 1.0. The dotted solid line represents the simulated values; the dashed line, the average of the values after time 20; and the solid line, the three standard error bounds around this proportion.

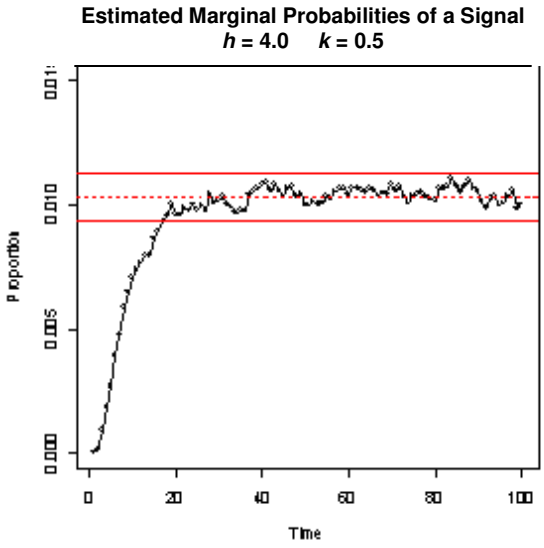


Figure 3.7: The estimated marginal probabilities of a signal for CUSUM parameter values of h = 4.0 and k = 0.5. The dotted solid line represents the simulated values; the dashed line, the average of the values after time 20; and the solid line, the three standard error bounds around this proportion.

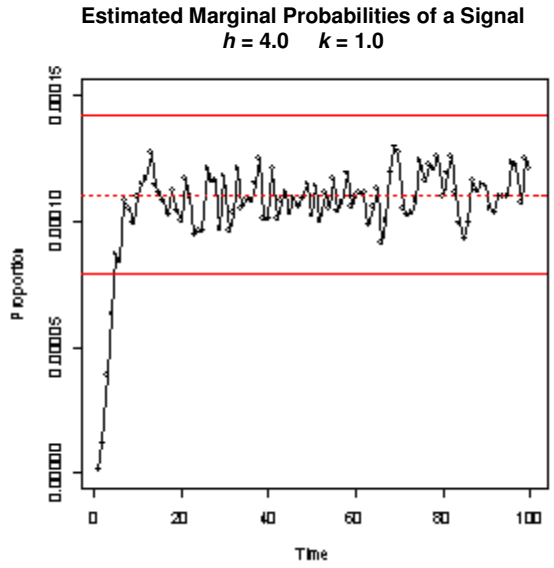


Figure 3.8: The estimated marginal probabilities of a signal for CUSUM parameter values of h = 4.0 and k = 1.0. The dotted solid line represents the simulated values; the dashed line, the average of the values after time 20; and the solid line, the three standard error bounds around this proportion.

3.3: The Recurrence Interval and Markov-Dependent Signaling Processes

A Markov-dependent signaling process is a model not a method used in practice as the EWMA chart or the CUSUM chart are (see Mousavi, 2006; for example). They are considered since they are reasonable and exact analytical results can easily be obtained. It should be understood that the Markov dependent signaling process is not a representation of the behavior of the EWMA or CUSUM chart. In the application of a Markov dependent signaling process presented here, the monitoring process is defined to be one for which there is a greater probability of a signal occurring after a signal has already occurred. A signal is more likely to follow a time period with a signal than to follow a time period with no signal. Suppose for such a Markov dependent signaling process the probability of a signal after a non-signal is a and the probability of a non-signal after a signal is b. Based on this, we have the following results:

1. The time-to-signal measure follows a geometric distribution with probability a where the time-to-signal is the number of time periods until the first signal. Therefore, the time-to-signal can take on any value 1, 2, 3, ….

2. The signal event is required to contain at least one signal. After this one signal occurs, the remaining length of the signal event follows a geometric distribution with probability b. Therefore, after the first signal in a signal event, the remaining length of the signal event can take on any value 0, 1, 2, ….

3. The time-between-signal events is required to contain at least one non-signal. After this one signal occurs, the remaining number of non-signals follows a geometric distribution

with probability a. Therefore, after the first non-signal, the remaining length of the time-between-signal events can take on any value 0, 1, 2, ….

Then, from these geometric distribution the in-control ATS = $1/a$, the ASEL = $1 + (1-b)/b = 1/b$, and the in-control ATBSE = $1 + (1-a)/a = 1/a$. The steady-state recurrence interval is obtained as the reciprocal of the long-run proportion of signals, $a/(a+b)$, and is, therefore, equal to $1 + b/a$. If it is impossible for signals to follow each other consecutively, i.e., if $b = 1$, then the steady-state recurrence interval is equal to the in-control ATS + 1, which is also the in-control ATBSE + 1.

As with the previously discussed monitoring methods, the recurrence interval is lacking in its usefulness in describing monitoring time-to-signal performance when applied to Markov-dependent processes. It is possible for processes with the same recurrence interval to have extremely different values of the in-control time-to-signal measures. Figures 3.9 and 3.10 show the behavior of the in-control time-to-signal measures when the recurrence interval is 101 and 1001, respectively. The in-control ATS, and equivalently the in-control ATBSE, shown in (a), increase as a decreases. The ASEL, shown in part (b) of these figures, increases as b decreases. The result is that these measures can take a wide range of values while the recurrence interval remains constant.

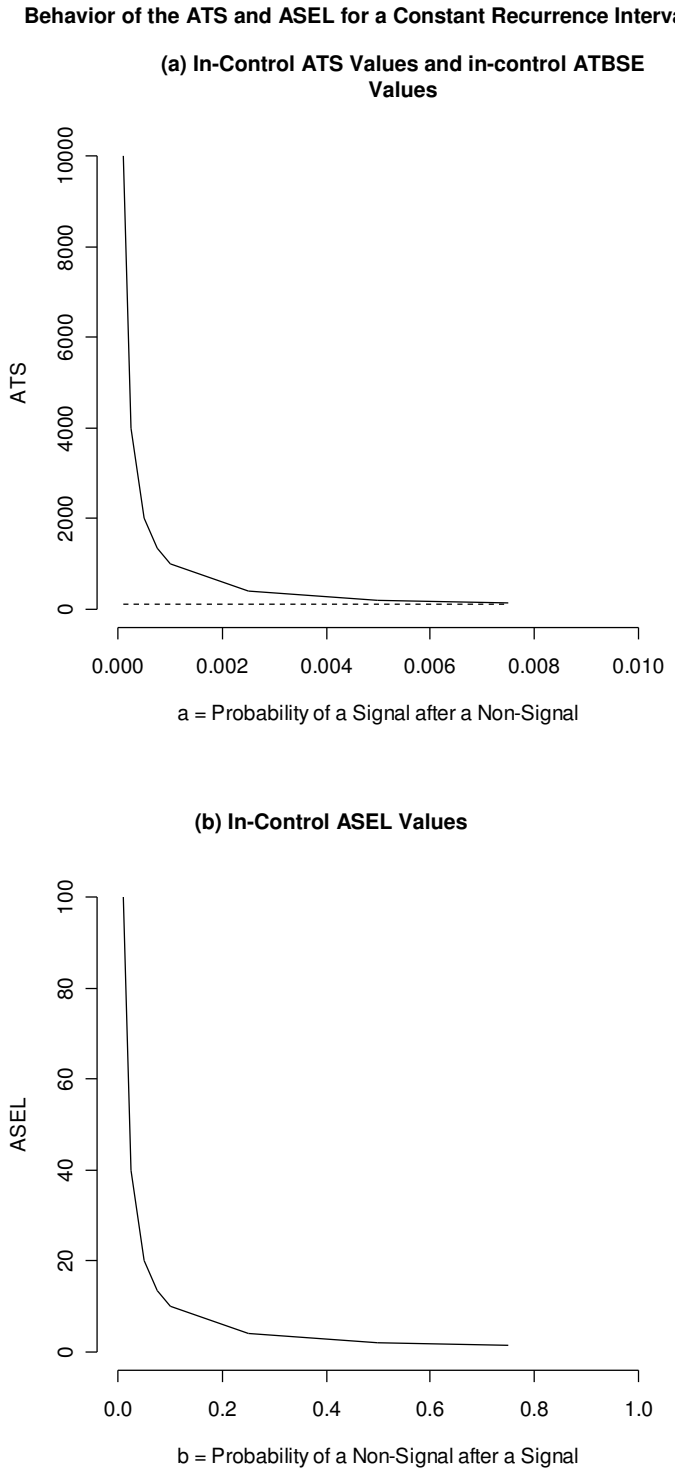


Figure 3.9: (a) Values for the in-control average time-to-signal and equivalently the in-control average time-between-signal events when the recurrence interval is constant at 101. The solid line represents the in-control ATS and the dashed line represents the recurrence interval. **(b)** Values for the in-control average signal event length when the recurrence interval is constant at 101.

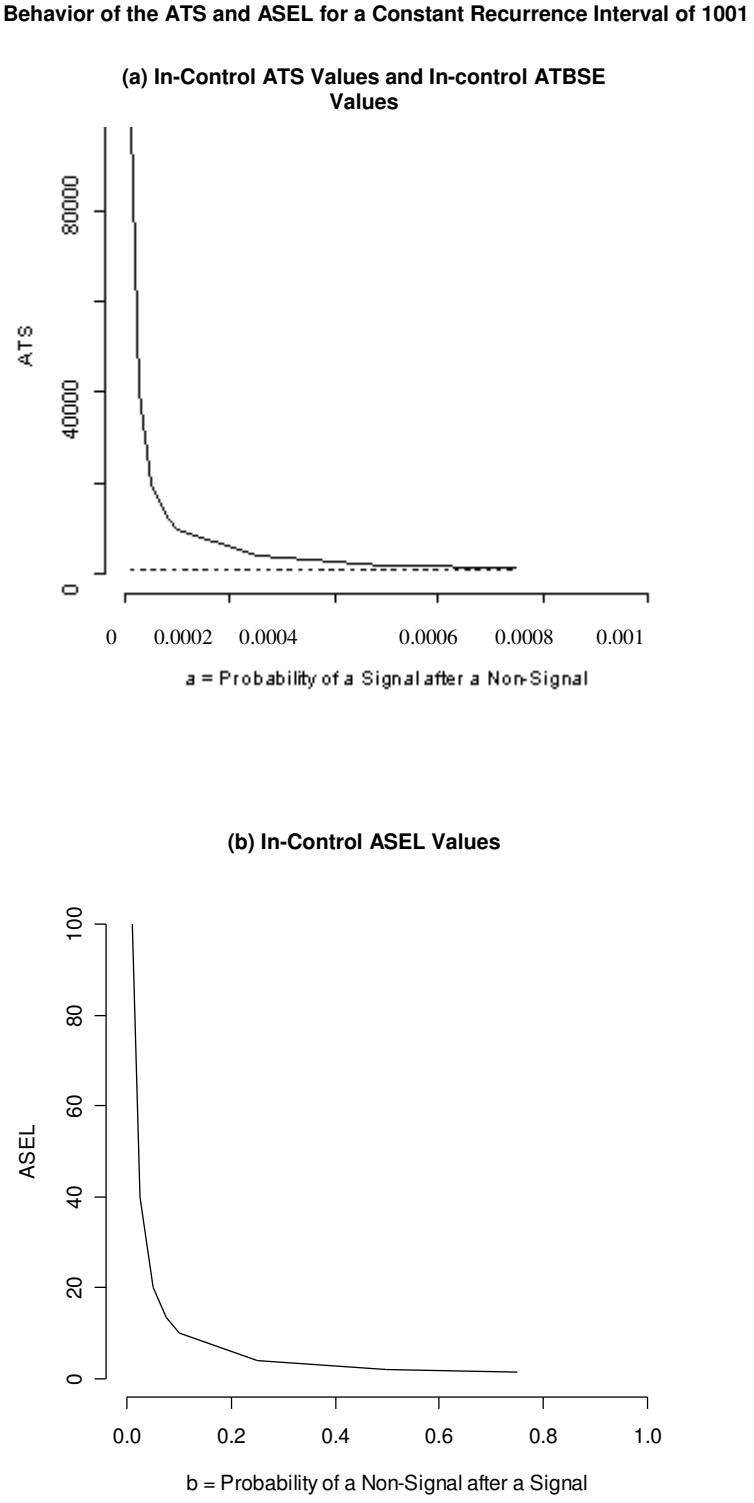


Figure 3.10: (a) Values for the in-control average time-to-signal and equivalently the in-control average time-between-signal events when the recurrence interval is constant at 1001. The solid line represents the in-control ATS and the dashed line represents the recurrence interval. **(b)** Values for the in-control average signal event length when the recurrence interval is constant at 1001.

The Markov-dependent signaling process also does not have constant marginal probabilities of a signal initially, similar to the case for the one-sided CUSUM chart. The initial marginal probabilities of a signal depend on the assumed starting state. The probability of a signal at the first time period is a if an initial state of no signal is assumed and $1 - b$ if an initial state of signal is assumed. A study was performed to see how quickly the Markov-dependent process reaches steady-state for different values of a and b for each of the two initial states. To do this, multiples of the two-by-two transition matrix, given as T below, were found

$$T = \begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix}. \tag{3.8}$$

The marginal probability of signal was then found by calculating

$$\underline{\pi}_0' T^{\,i}, \tag{3.9}$$

for $i = 0, 1, 2, \ldots$; where $\underline{\pi}_0' = (1,0)$ when the initial state corresponds to a non-signal and $\underline{\pi}_0' = (0,1)$ when the initial state corresponds to a signal. Bhat and Lal (1990) showed that the limiting probabilities

$$\pi_j = \lim_{n\to\infty} t_{ij}^{(n)} \text{ where } t_{ij} = P(X_n = j \mid X_o = i) \tag{3.10}$$

exist independent of the initial state as long as $|1 - a - b| < 1$. For the Markov dependent signaling processes examined here, this condition will always be met. Therefore, once steady-

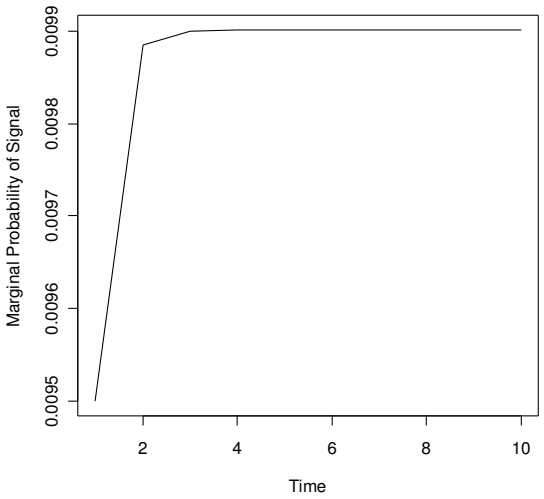
state is reached, the marginal probabilities of a signal will be constant and the steady-state recurrence interval will be equal to the reciprocal of this value, $a/(a+b)$, regardless of the initial state of the process.

Figures 3.11 and 3.12 show the marginal probabilities of a signal for a Markov dependent signaling process with (a) an initial state of non-signal and (b) an initial state of signal. For a = 0.0095 and b = 0.95, steady-state seems to be reached around time period five, regardless of the initial state of the process. The marginal probability of a signal in steady-state is 0.0099 for either initial state. For a = 0.0075 and b = 0.75, steady-state seems to be reached around time period six if it is assumed that the initial state of the process is a non-signal, while nine time periods are required to reach steady-state if it is assumed that the initial state of the process is a signal. Again, the marginal probability of a signal in steady-state is 0.0099.

Marginal Probabilities of a Signal for a Markov Process

***a* = 0.0095 *b* = 0.95**

(a) Initial State of Non-Signal



(b) Initial State of Signal

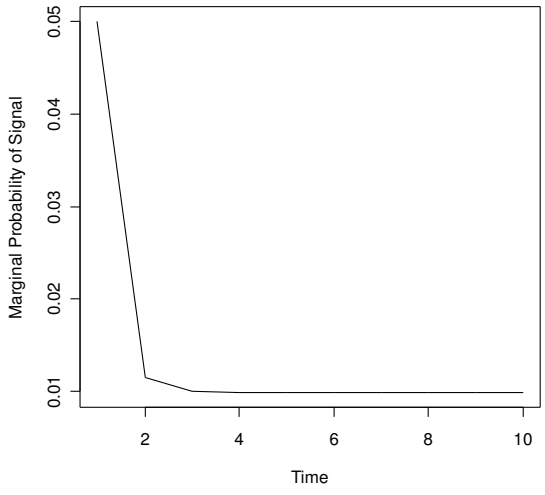
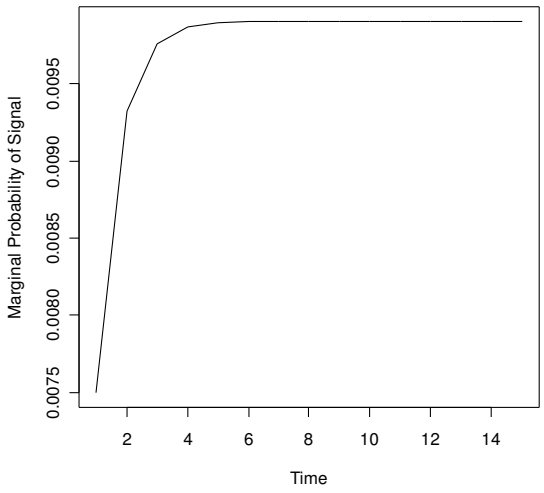


Figure 3.11: (a) Plot of the marginal probabilities of signal for a Markov dependent process with $a = 0.0095$ and $b = 0.95$ when the initial state corresponds to a non-signal. **(b)** Plot of the marginal probabilities of signal for a Markov dependent process with $a = 0.0095$ and $b = 0.95$ when the initial state corresponds to a signal.

Marginal Probabilities of a Signal for a Markov Process
a = 0.0075 b = 0.75

(a) Initial State of Non-Signal



(b) Initial State of Signal

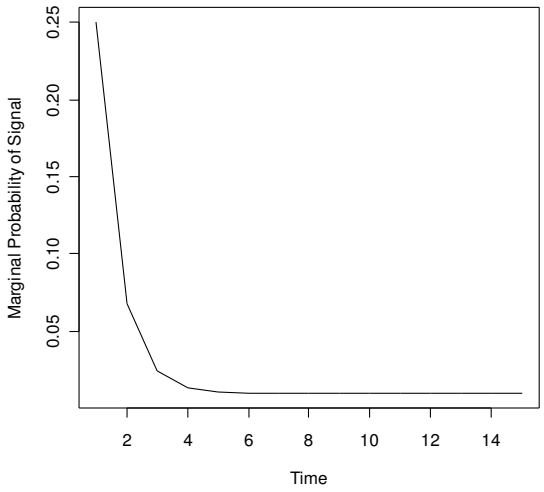


Figure 3.12: (a) Plot of the marginal probabilities of signal for a Markov dependent process with a = 0.0075 and b = 0.75 when the initial state corresponds to a non-signal. **(b)** Plot of the marginal probabilities of signal for a Markov dependent process with a = 0.0075 and b = 0.75 when the initial state corresponds to a signal.

Table 3.4 shows the amount of time needed until the absolute difference in consecutive marginal probabilities of a signal is less than 0.00001 for different values of a and b for the two different initial conditions. As b increases, the amount of time needed to reach steady-state decreases. Also, more time is needed to reach a steady-state when the initial condition is assumed to correspond to a signal. For either initial case, if $a + b = 1$ then the convergence is immediate and the Markov dependent signaling process would not be necessary.

a	$1 - b$	*Time to Reach Steady-State Initial State of Non-Signal*	*Time to Reach Steady-State Initial State of Signal*
0.0095	0.05	5	5
0.0025	0.05	3	5
0.0085	0.15	5	7
0.0025	0.15	4	7
0.0075	0.25	6	9
0.0025	0.25	5	10
0.0050	0.50	10	17
0.0025	0.50	9	17

Table 3.4: Time needed to reach steady-state in a Markov dependent process depending on the initial state for different values of a and b.

3.4: Discussion

The time-to-signal metrics and the recurrence interval do not provide similar information about a surveillance process though they are commonly confused in the literature. Both measurements, however, are used extensively; the recurrence interval in public health

surveillance and time-to-signal measures in industrial SPC. The in-control ATBSE was introduced as a measure to compare time-to-signal properties with the recurrence interval. The introduction of such a measure is necessary because the in-control ATS only summarizes time-to-signal performance with respect to the first signal event, whereas the recurrence interval can be used at any phase in the surveillance process as long as the process remains in-control. The in-control ASEL was also introduced for surveillance schemes that are not being reinitialized after the first signal. Consecutive signals, on the same side of the centerline for two-sided charts, are being treated as one signal event. By estimating the in-control ATBSE, we show that if an EWMA chart is not reset after the first signal event, it is more likely to signal again sooner than a control chart at its initial settings. This result is intuitive and would be expected to hold for all except basic Shewhart-type Charts.

The time-to-signal properties seem to be much more useful than the recurrence interval. The recurrence interval must be defined in terms of a steady-state measure when the $P(E_i = 1)$ are not constant. The amount of time required to reach steady-state, however, is dependent on the parameters of the control chart being employed. More importantly, charts with the same recurrence interval value can have quite different time-to-signal properties. Also, regardless of how the recurrence interval is defined, in terms of a steady-state measure where applicable or not, some time-to-signal measures are still required to assess out-of-control performance since the recurrence interval is only defined for in-control processes.

The use of the ATBSE and ASEL in the out-of-control case is dependent on the method of surveillance. For CUSUM control charts used in industrial SPC, if the process is not adjusted

after a signal, the ATBSE will approach zero and the ASEL will approach infinity since the monitoring statistic will likely never return to a non-signal state. This, however, is not the case for Shewhart charts and not always true for EWMA charts. It also might not be true for other surveillance techniques, as will be discussed in Chapter 4.

Chapter 4

Evaluation of Kulldorff's Scan Method in the Temporal Setting

Scan statistics remain the basis of many popular methods employed in disease monitoring. With the availability of the free software *SatScan*™, the scan method proposed by Kulldorff in 1997 and then extended by him in 2001 is commonly applied in disease surveillance. The properties of this method in the prospective case, however, have not been studied carefully. In this chapter, the scan statistic developed by Kulldorff (1997 and 2001) will be evaluated in the temporal scenario where a Poisson model is assumed and incidences occur independently over time.

4.1: The Temporal Scan Statistic

Temporal data contain the time of incidences reported either at their exact time of occurrence or aggregated into non-overlapping intervals such as days, weeks, or years. No spatial information is available. Throughout this chapter and the next, it is assumed that the data are aggregated into equally sized time intervals.

Wallenstein and Naus (2004) and Naus and Wallenstein (2006) derived a scan statistic to be used in the temporal case. An increase in the incidence rate is signaled if the number of occurrences of the event in an interval of time of fixed length is significantly larger than expected. Joner et al. (2007) examined the properties of this method in the Bernoulli case using Markov chains. They showed that Bernoulli CUSUM charts of Reynolds and Stoumbos (1999) have better performance than the scan statistic of Naus and Wallenstein (2006) under a sustained shift in the incidence rate. The two methods seem to be roughly equivalent in their ability to detect an increased rate under the assumption of a shift in the incidence rate of limited duration.

Another popular scan statistic was proposed by Kulldorff (1997 and 2001). One difference between this method and the one proposed by Naus and Wallenstein is that the interval of time is no longer fixed. Intervals of any size, with some restrictions explained later in this section, are considered as possible time periods of an increased incidence rate. The scan method as explained by Kulldorff (1997) can be applied to Bernoulli or Poisson count data. The log-likelihood ratio (LLR) statistic is used. For the Poisson model, this can be found by taking the natural log of Equation (2.8). The result is

$$LLR = \ln\left(\frac{L(x)}{L_o}\right) = x\ln\left(\frac{x}{E_s}\right) + (N-x)\ln\left(\frac{N-x}{N-E_s}\right) \quad (4.1)$$

if $x > E_s$ and $ln(L(x)/L_0) = 0$ otherwise. Since this surveillance statistic is being applied to the temporal case, the terms in the equation have a slightly different meaning than they do in the spatial case. Now, N is the total number of observed incidences for all time periods, x is the observed number of incidences in the sub-interval of time being considered as a possible cluster, and E_s is the expected number of incidences for that sub-interval. Because this method conditions on the total number of observed incidences, N, the expected rate for any one unit of time is estimated from the data by the number of incidences that have occurred up until the current time period of surveillance divided by the current number of time periods of observation. The value of E_s is, therefore, found as this estimated rate times the length of the sub-interval under consideration. In the prospective monitoring case, this means with each new additional time period observed, the estimated incidence rate will most likely change.

To determine what, if any, sub-interval(s) is a significant time of increased incidence rate, the LLR is calculated for all possible interval lengths. The relevant intervals in the prospective case must end at the current time period. The maximum LLR over all of these values is the observed surveillance statisitc, hereafter referred to as the maximum observed LLR. A large value of the maximum observed LLR would imply that the sub-interval is a possible cluster corresponding to a significantly higher incidence rate.

To determine what value of the maximum observed LLR is large enough to imply an increase in the incidence rate, the distribution of the maximum LLR must be estimated using

Monte Carlo (MC) simulation. Because the LLR calculation for this method is conditioned on the total number of observations, the observed number of incidences until the current time period are distributed across the current number of time periods using the multinomial distribution with the probability of an incidence occurring in each interval being equal. Adjustments for various outside influences, such as the population at that time, can be made during the MC simulation by using probabilities of incidences occurring at each time period that are not equal.

This process of distributing the observations across the time periods is done repeatedly providing multiple alternative datasets that would result in the same number of incidences in the same number of recorded time periods. The maximum LLR is found for each possible interval length, i.e. the largest LLR for all intervals of length one is found, the largest LLR for all intervals of length two is found, etc., for all of the MC replications. The maximum LLR of a MC replication is found as the maximum over all possible interval lengths. Taking the overall largest LLR for each MC replication results in an estimated distribution of the maximum LLR value. The rank of the maximum observed LLR is then found in relation to this distribution. According to the *SaTScan*™ software manual (Kulldorff 2005a), "if this rank is *R*, then $p = R / (1 + \#\ simulation)$" where "*# simulations*" is the number of MC replications and "*p*" is the *p*-value of the maximum observed LLR. Kulldorff (1997), alternatively, stated that "with 9999 such replications [of the MC], the test statistics is significant at the 5 percent level if the value of the test statistics for the real data is among the 500 highest values of the test statistic coming from the replications".

There is an issue, however, with these two statements. It is possible for the surveillance statistic for the observed data to be tied with values of surveillance statistics from the MC simulated data. This will be especially likely for earlier time periods. For instance, suppose there are 10 total incidences observed in two time periods. During the MC simulation, there are only 11 possible distributions of these 10 incidences across the two time periods; these are 0 incidences at the first time and 10 at the second time, 1 incidence at the first time and 9 at the second time, etc. If there are more than 11 simulated data sets generated in the MC then there have to be repeated values. The *p*-value definitions from Kulldorff (2005a) and Kulldorff (1997) do not address this issue of ties. Based on the Kulldorff (1997) definition, there is no guarantee that the 500^{th} ranked value of the simulated surveillance statistic is different from the 501^{st} ranked value of the simulated surveillance statistic, or in an extreme case, even the 600^{th} ranked value or the 50^{th} ranked value, for example. The *p*-value could be greatly altered by this fact. Using the definition from Kulldorff (2005a), there are at least three different ways the rank could be assigned to the observed surveillance statistic in the presence of ties; these include, but are not limited to, giving the maximum observed LLR the minimum rank, the maximum rank, or the average rank of all tied values. Here, the average rank of all tied maximum LLR values will be given to the observed surveillance statistic in the event of ties.

A second issue is also present in the calculation of the surveillance statistic both from the observed data and in the MC simulation. If the observed number of incidences in a sub-interval of time is the same at the total number of incidences that occurred, then the LLR value is undefined. This happens because $N - x$ would be zero and the second term in the LLR would be *0ln(0)*. This is guaranteed to occur when the entire monitoring period is the "sub-interval" of

interest but could also occur at other times. This issue does not appear in the spatial or spatiotemporal cases because the circles or cylinders, as appropriate, are restricted to contain no more than half of the observed incidences. In the temporal case, Kulldorff (2005a) stated "For purely temporal and space-time analysis, the maximum temporal cluster size can be specified in terms of a percentage of the study periods as a whole or as a certain number of days, months or years. The maximum must be at least as large as the length of aggregated time interval length. If specified as a percent, then for the Bernoulli and Poisson models, it can be at most 90 percent, and for the space-time permutation model, at most 50 percent. The recommended value is 50 percent."

This restriction attempts to correct the problem where a sub-interval of time would contain all N observations. The issue is when the incidence rate is small and there is the possibility of a number of time periods in which zero incidences were observed. For instance, if the observed number of incidences in 10 time periods were as follows

0 0 0 0 0 1 1 0 0 0

then at time $T = 10$ there are $N = 2$ total incidences. Restricting a sub-interval size to be no more than 50% of the total time periods results in at most 5 time periods in any possible cluster. In the prospective case, the largest interval would then be the one that contained the incidences {1, 1, 0, 0, 0}. Therefore, the number of observed incidences in the sub-interval is still equal to the total number of incidences, N.

Kulldorff (2001) proposed an adjustment to his scan statistic *p*-values for the prospective case. This adjustment is recommended to correct, from a multiple testing point of view, for all previous tests of possible clusters. The clusters examined in the adjusted MC simulation are still of varying size but are allowed to occur at any point in the surveillance process. This means the sub-intervals are no longer restricted to end at the current time period during the estimation of the distribution of the maximum LLR. For example, when searching for the cluster of size one time interval that has the largest LLR in the simulation, one could find it to be the first time period of surveillance even if the current time period is the fiftieth. In the "unadjusted" case, this could not happen. In this situation, the sub-intervals are required to end at the current time period as in the prospective case. Under both situations, the maximum observed LLR will be identical since any observed cluster must end at the current time period in prospective monitoring. Because of this, the adjusted method will always result in a *p*-value that is no less than the *p*-value calculated by the unadjusted method.

Lastly, since the MC simulation is a vital component to this scan method, it is important to understand its effect on the outcome. At different times, Kulldorff reports different values for the number of MC replications, *S*, used. For instance, sometimes 999 MC replications are performed while other times 9999 MC replications are carried out. In either case, the number of MC replications is one less than a power of ten to simplify the ranking method. That is, using 999 MC replications will result in the *p*-values calculated being equal to the rank of the maximum observed LLR divided by 1,000. Similarly, with 9999 MC replications, the *p*-values will be the rank of the maximum observed LLR divided by 10,000.

Figure 4.1 shows the estimated distribution of the maximum LLR from the MC replicates for a selected number of time periods for the case where the limit on the interval size is 50% of the total number of time periods. A rate of 5 incidences per time interval is assumed when generating the data. The distribution of the maximum LLR in both the adjusted and unadjusted cases seems to be highly skewed. For this reason, both 999 and 9999 MC replicates will be examined to see if this number has a significant effect on the results of the overall simulation.

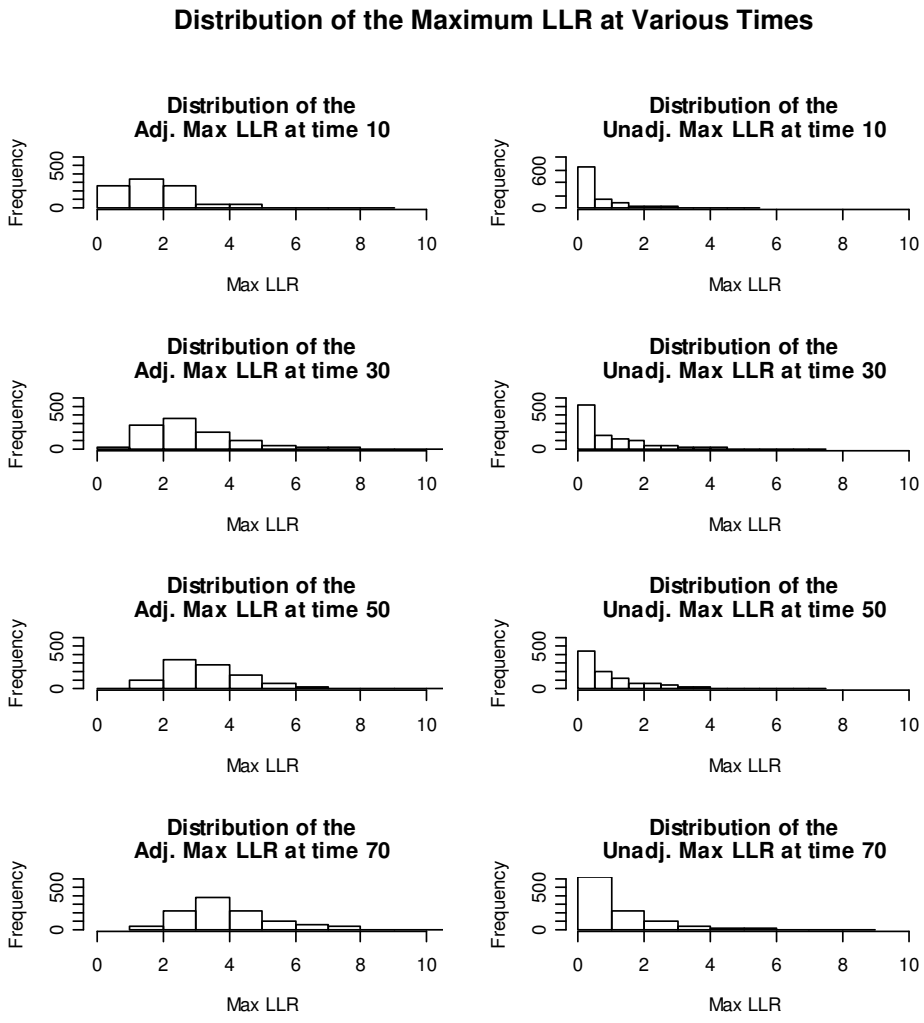


Figure 4.1: Distribution of the maximum log-likelihood ratio surveillance statistic at various times using both the adjusted and unadjusted methods for the restricted case where a cluster can not contain any more than 50 percent of all time periods.

4.2: Evaluation Under a Stable Incidence Rate

To obtain the results in this section, data were generated under the assumption that there is no increase in the incidence rate. The maximum observed LLR was based on simulated data that follow a Poisson distribution with some assumed rate. The incidence rate was set to be 5, 10, or 50 incidences per time period. The temporal data were aggregated into non-overlapping intervals of fixed size with 250 total time intervals simulated. Each dataset was examined using the Kulldorff scan method. The surveillance statistic was calculated for the prospective case and, therefore, each cluster considered must end at the current time period. The maximum size of the possible sub-intervals of time was set to be 50%, 75%, or 90% of the total monitoring length. Both the adjusted and unadjusted *p*-values are calculated. For all combinations of the incidence rates, number of MC replications, and percentages, 1,000 such datasets were generated.

Figure 4.2 shows the observed incidences, the adjusted *p*-values and unadjusted *p*-values for one of the 1,000 iterations performed for the scenario when $\mu = 10$, percent = 75%, and $S = 999$. The fact that the unadjusted *p*-values will always be at least as small as the adjusted *p*-values can clearly be seen by comparing the middle and lower plots. There are a few spiked decreases in the adjusted *p*-values, which correspond to the slightly higher than expected number of incidences. These are not large, however. There are a few decreases over time in the unadjusted *p*-values. Here, there are five significant signal events using $\alpha = 0.2$. Figure 4.3 shows the autocorrelation function for both the adjusted and unadjusted *p*-values from Figure 4.2. There is a significantly high degree of positive autocorrelation in these *p*-values. As will be shown, this will mean that the recurrence interval metric is not related to the time-to-signal performance of the scan method.

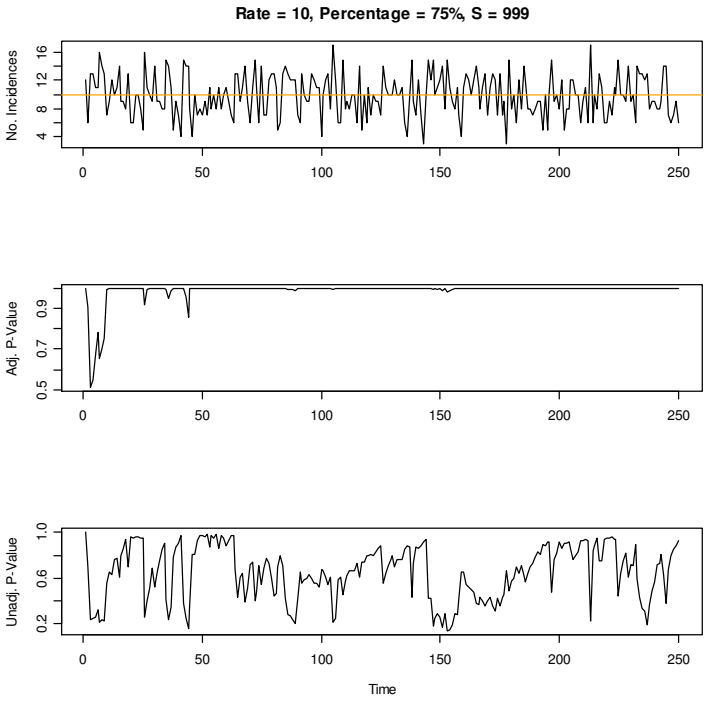


Figure 4.2: An example of the number of incidences, adjusted *p*-values, and unadjusted *p*- values for the scenario when $\mu = 10$, percentage = 75%, and $S = 999$.

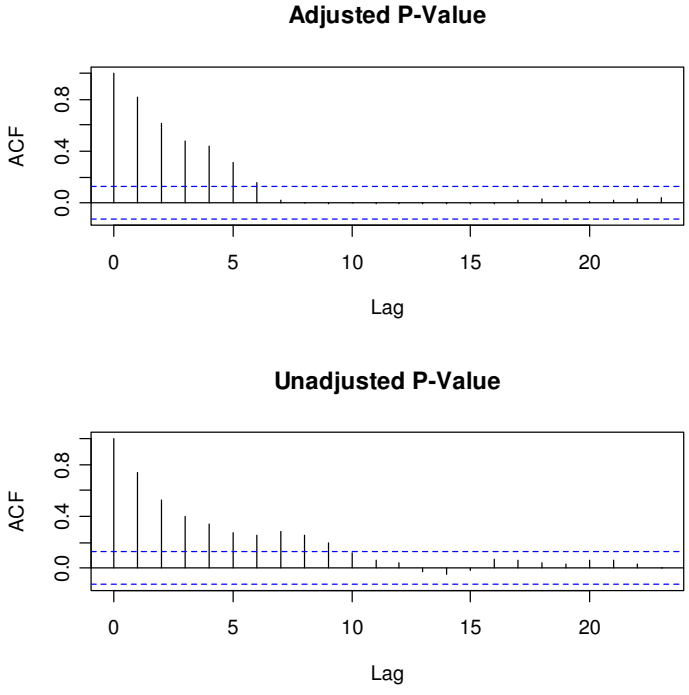


Figure 4.3: Autocorrelation function for the adjusted *p*-values and unadjusted *p*- values for the scenario when $\mu = 10$, percentage = 75%, and $S = 999$ from Figure 4.2.

Figure 4.4 and Figure 4.5 show a second iteration from the same simulation. Here, however, there is quite a long signal event even though there is no increase in the incidence rate. The dips in the plots of the *p*-values correspond to time periods in which the observed number of incidences is larger then the estimated incidence rate. Any of the time periods in which the *p*-value falls below the cut-off value, α, would be considered a significant increase in the incidence rate. In Figure 4.2, there is a large dip in both *p*-values starting around time period 30 and continuing until after time period 50. This corresponds to the same consecutive time periods in the plot of the number of incidences that all fall above the estimated incidence rate. After this one section of decreases in the *p*-values, the adjusted *p*-value remains large for the remaining time periods of surveillance with only occasional very small decreases. The unadjusted *p*-values, however, take more time to recover from the false signal event and vary more widely.

Part of the reason for the large magnitude of the decreases in the *p*-values early in the monitoring process is due to the fact that the estimation of the incidence rate is more unstable at the beginning of the process. This means that since there are less time periods observed and, therefore, less information available for the estimation of the incidence rate, slight variations in the number of observed incidences will have a greater influence on the estimated incidence rate. This will tend to result in decreases in the *p*-values at earlier time periods. Once more time has passed in the monitoring process, the estimated incidence rate is not as influenced by a few extreme observations and the decreases in the *p*-values will tend to be smaller.

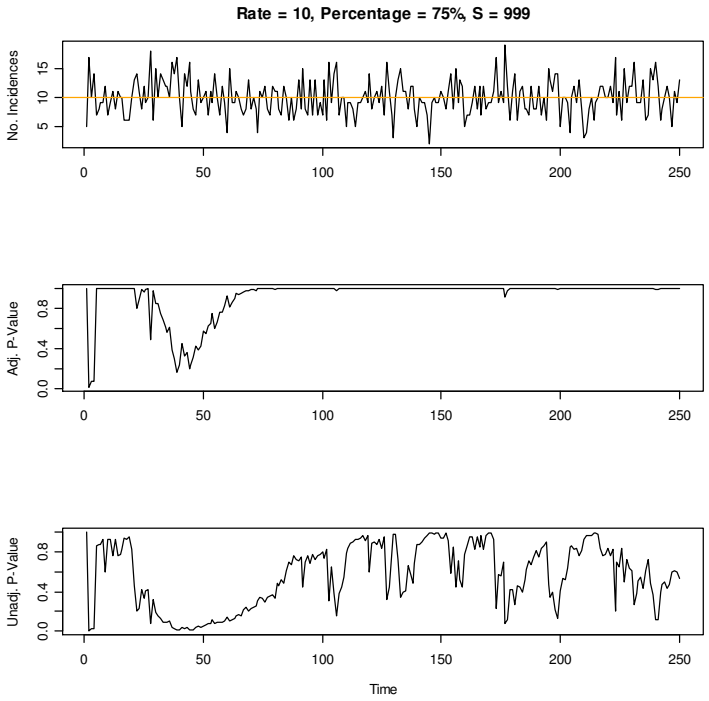


Figure 4.4: An example of the number of incidences, adjusted *p*-values, and unadjusted *p*- values for the scenario when $\mu = 10$, percentage = 75%, and $S = 999$.

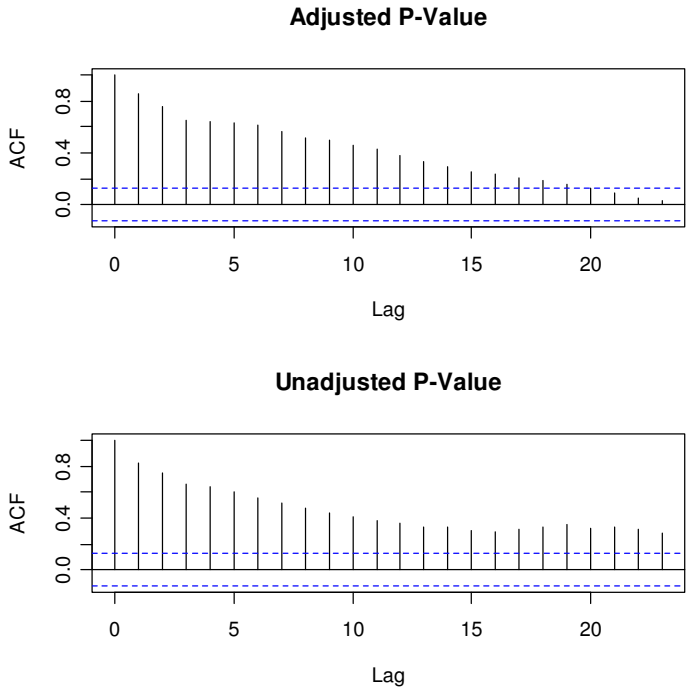


Figure 4.5: Autocorrelation function for the adjusted *p*-values and unadjusted *p*- values for the scenario when $\mu = 10$, percentage = 75%, and $S = 999$ from Figure 4.4.

According to Kleinman (2005), the use of the recurrence interval is preferable to the adjusting of *p*-values for multiple testing over time. Kulldorff (2005b) agreed with this assessment. As stated previously, however, the recurrence interval can only be applied when the process being monitored is in-control and when the marginal probabilities of a signal are constant. Throughout this section, it is assumed the process is in-control, therefore the marginal probabilities of a signal need to be examined to see if they are indeed constant. To do this, similar methods to the CUSUM analysis shown in Section 2 of Chapter 3 were performed. The proportions of signals that occur at each time period are plotted against time in Figure 4.6 and 4.7 for different μ and percentage with S = 999. In both situations, the marginal probability of a signal seems relatively constant for the unadjusted *p*-value approach but not for the adjusted *p*-value approach. The adjusted *p*-value approach is more likely to result in a signal early in the process than it is later in the process. For this reason, the recurrence interval can be applied to the situation where no adjusting for previous tests is done, but not for the situation where adjusting is performed. It should be noted that the scan method proposed by Kulldorff will never signal at the first time period, as seen in both plots. This is due to the fact that the observed number of incidences at the first time period will always be the same as the total number of observations and therefore, the test can not be performed. The adjusted and unadjusted *p*-values were both arbitrarily set to one in this situation which results in the marginal probabilities of a signal being zero at time period one.

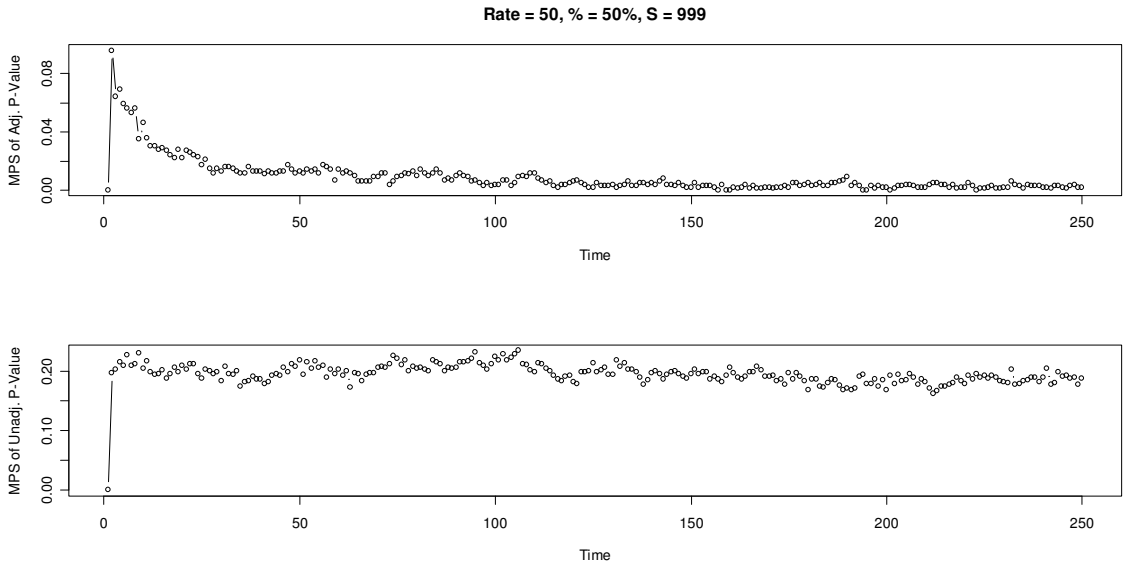


Figure 4.6: The estimated marginal probabilities of a signal for the adjusted and unadjusted p-values when μ = 50, percentage = 50%, and S = 999.

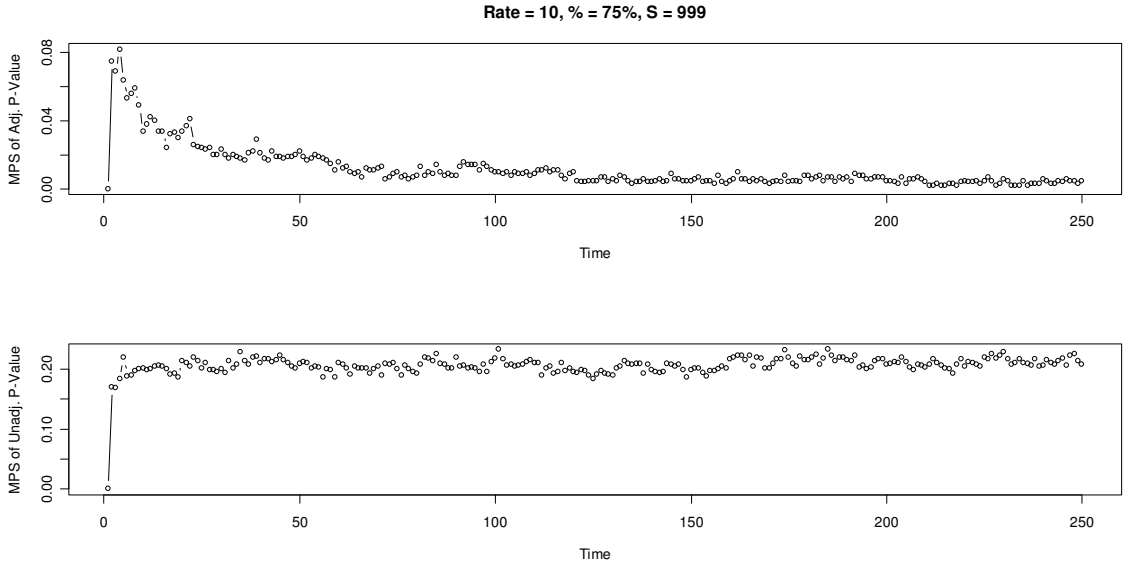


Figure 4.7: The estimated marginal probabilities of a signal for the adjusted and unadjusted p-values when μ = 50, percentage = 50%, and S = 999.

The in-control ATS, in-control ATBSE, in-control ASEL and recurrence interval were estimated for each combination of assumed incidence rate, μ, number of MC replications, S, and percentage of time periods allowed to be considered a cluster, %, for the unadjusted p-values. Also calculated are the mean, median, minimum, and maximum value for both the adjusted and unadjusted p-values, calculated over all replications. These results are given in Table 4.1 along with the standard errors. The recurrence interval errors are reported in terms of a confidence interval because the marginal probability of a signal was estimated from the results of the simulation and a confidence interval for this probability found. The estimate for the recurrence interval is then the reciprocal of the estimated marginal probability of a signal and the confidence interval limits on the recurrence interval are the reciprocals of the confidence interval bounds for the marginal probability of a signal. The reason that the in-control ATS, in-control ATBSE, in-control ASEL, and recurrence interval are not available for the adjusted p-value approach is because these p-values are extremely large. A cut-off value of $\alpha = 0.2$ was used for the unadjusted case so that all iterations result in at least one signal. For this to be possible for the adjusted case, the cut-off value would have to be extremely large.

The maximum p-value over all iterations for both the adjusted and unadjusted p-values is always 1.0. The minimum p-value over all iterations for both the adjusted and unadjusted p-values is always $(S + 1)^{-1}$ except in the case of $\mu = 50$, $S = 9999$, and percentage = 50%. The fact that the maximum p-value is equal to 1.0 and the minimum p-value is equal to $(S + 1)^{-1}$ is not surprising due to the way the p-values are calculated. A p-value = 1.0 means that the maximum observed LLR was the last in the ordered list and therefore received a rank of $S + 1$, without being tied with another observation. A p-value = $(S + 1)^{-1}$ means that the maximum observed

LLR was the first in the ordered list and therefore received a rank of 1, without being tied with another observation. The mean and median for the unadjusted *p*-values are both close to 0.5. The medians of the adjusted *p*-values are always 1.0 and the means of the adjusted *p*-values are around 0.95, again an indication that the adjusted *p*-value is always larger than the unadjusted *p*-value. It seems that the method advocated by Kulldorff (2001) for correcting the multiple testing over time is overcorrecting and that the over adjustment increases over time.

			Unadjusted *p*-value Approach								Adjusted *p*-value Approach			
μ	***S***	***%***	***ATS***	***ATBSE***	***ASEL***	***RI***	***mean***	***median***	***min***	***max***	***mean***	***median***	***min***	***max***
5	**999**	**50**	8.935	15.712	4.234	5.03	0.5021	0.503	0.001	1	0.9545	1	0.001	1
			(0.4091)	(0.1886)	(0.0868)	(4.22,6.20)	(0.0032)				(0.0013)			
5	**999**	**90**	15.762	19.762	5.684	5.01	0.5002	0.501	0.001	1	0.9360	1	0.001	1
			(0.7566)	(0.2968)	(0.1863)	(4.21,6.18)	(0.0032)				(0.0015)			
10	**999**	**50**	9.126	15.357	4.297	4.85	0.4999	0.4998	0.001	1	0.9554	1	0.001	1
			(0.3913)	(0.1821)	(0.0866)	(4.09,5.96)	(0.0031)				(0.0013)			
10	**999**	**75**	13.234	17.454	5.049	4.86	0.4991	0.498	0.001	1	0.9455	1	0.001	1
			(0.5510)	(0.2357)	(0.1337)	(4.10,5.97)	(0.0032)				(0.0013)			
10	**999**	**90**	16.524	19.038	5.665	4.87	0.4989	0.4965	0.001	1	0.9383	1	0.001	1
			(0.7175)	(0.2806)	(0.1784)	(4.11,5.99)	(0.0032)				(0.0015)			
50	**999**	**50**	8.289	15.366	4.024	5.12	0.5055	0.508	0.001	1	0.9579	1	0.001	1
			(0.3559)	(0.1785)	(0.0794)	(4.30,6.34)	(0.0032)				(0.0013)			
50	**999**	**90**	14.415	18.996	5.345	5.07	0.5061	0.5115	0.001	1	0.9410	1	0.001	1
			(0.6772)	(0.2724)	(0.1719)	(4.25,6.27)	(0.0032)				(0.0015)			
50	**1999**	**50**	8.157	15.489	4.061	5.11	0.5053	0.5078	0.0005	1	0.9579	1	0.0005	1
			(0.3452)	(0.1799)	(0.0810)	(4.29,6.33)	(0.0032)				(0.0013)			
5	**9999**	**50**	8.730	15.812	4.280	5.01	0.5016	0.5032	0.0001	1	0.9545	1	0.0001	1
			(0.4016)	(0.1903)	(0.0879)	(4.21,6.18)	(0.0032)				(0.0013)			
50	**9999**	**50**	8.286	15.601	4.095	5.11	0.5051	0.5077	0.0001	1	0.9579	1	0.00035	1
			(0.3501)	(0.1813)	(0.0819)	(4.29,6.33)	(0.0033)				(0.0013)			

Table 4.1: Performance metric summary for the in-control scan method. In-control ATS, in-control ATBSE, in-control ASEL, and recurrence interval for the unadjusted *p*-values, as well as the mean, median, minimum, and maximum values for both the unadjusted and adjusted *p*-values for different values of the number of incidences, MC replications, and maximum sub-interval size. For all combinations, there are T = 250 total time periods and $\alpha = 0.2$. Standard error values are given in parentheses.

In Figure 4.8, the effect the change of S has on the in-control ATS, in-control ATBSE, and in-control ASEL is examined. The same data were examined with the Kulldorff scan method calculated with the number of MC replications equal to 999 and then again with the number of MC replicates equal to 1999 and, lastly, with the number of MC replicates equal to 9999. Therefore, the only factor that has changed is the number of MC replications used to estimate the distribution of the maximum LLR. The solid line represents the estimated values of these lines while the dotted lines represent the three standard error bounds around these estimates. A change in S does not seem to have a significant effect on these estimates. Therefore, using S = 999 is recommended due to the fact that it requires less computational time.

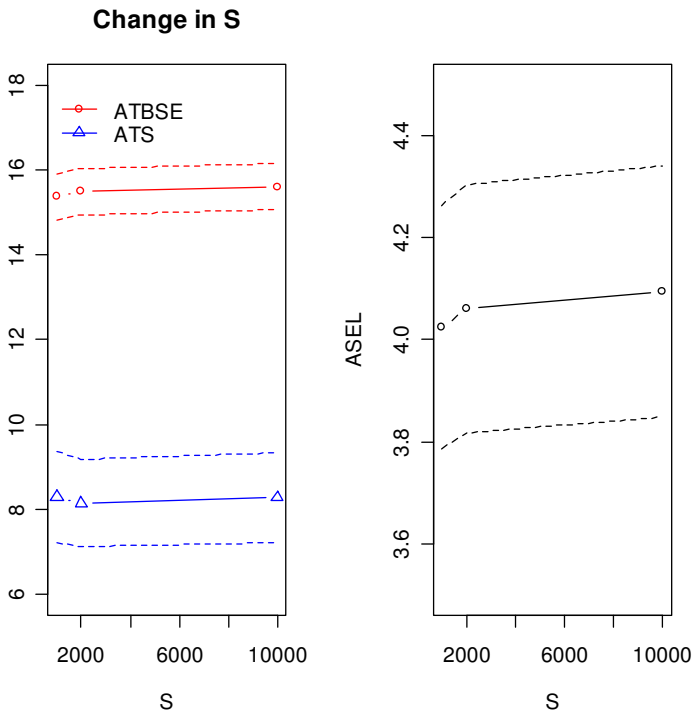


Figure 4.8: Effect of a change in the number of MC replications, S, on the in-control time-to-signal properties of the Kulldorff scan method. The rate is assumed to be $\mu = 50$.

Figure 4.9 shows the effect that a change in the Poisson rate has on the in-control time-to-signal properties of the Kulldorff scan method. The number of MC replications is $S = 999$. Again the solid line represents the estimated value and the dotted lines are the three standard error bounds around this estimate. These values were calculated for both a percentage of 50%, in Figure 4.4 (a), and 90%, in Figure 4.4 (b). Again, the rate does not seem to have a significant effect on these measures.

Figure 4.10 examines the change in the in-control time-to-signal properties as the percentage of the time periods allowed to be in a cluster changes. The same data were examined with the Kulldorff scan method calculated with percentages of 50%, 75% and 90%. Therefore, the only factor that has changed in the size of the possible sub-intervals of time. Again the solid line represents the estimated value of the metrics and the dotted lines are the three standard error bounds around these estimates. The incidence rate is assumed to be $\mu = 10$. It appears that the percentage does have a significant effect on these measures. As the maximum size of a sub-interval increases, the in-control time-to-signal measures also increase. Typically, it is desirable to have a larger in-control ATS and in-control ATBSE. This would mean that the process results in fewer false alarms. Based on this, a maximum interval size of 90% would seem to be the best choice of maximum cluster size. Using a maximum size of 90% of time, however, does mean that 90% of all time periods are included as anomalies while 10% of the time periods are not.

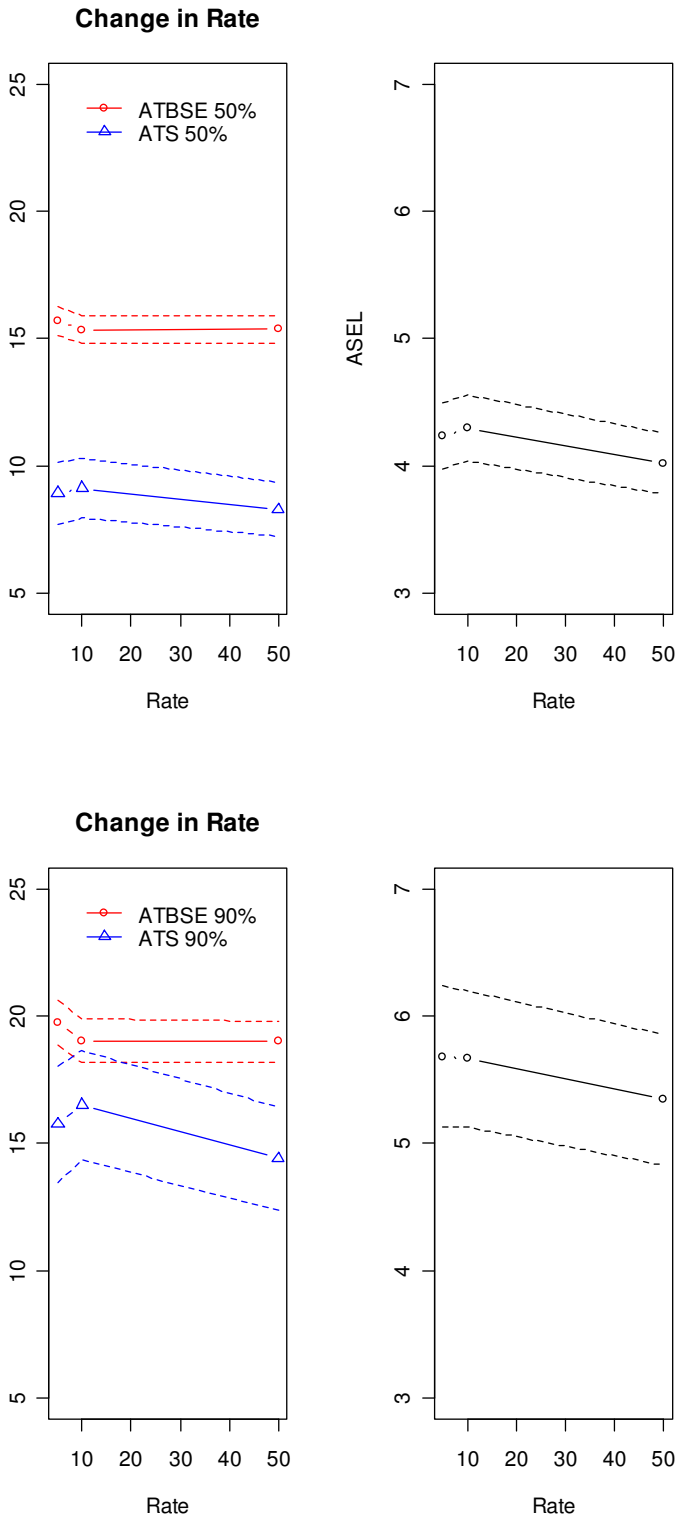


Figure 4.9: Effect of a change in the incidence rate, μ, on the in-control time-to-signal properties of the Kulldorff scan method. The percentage is assumed to be (a) 50% or (b) 90%.

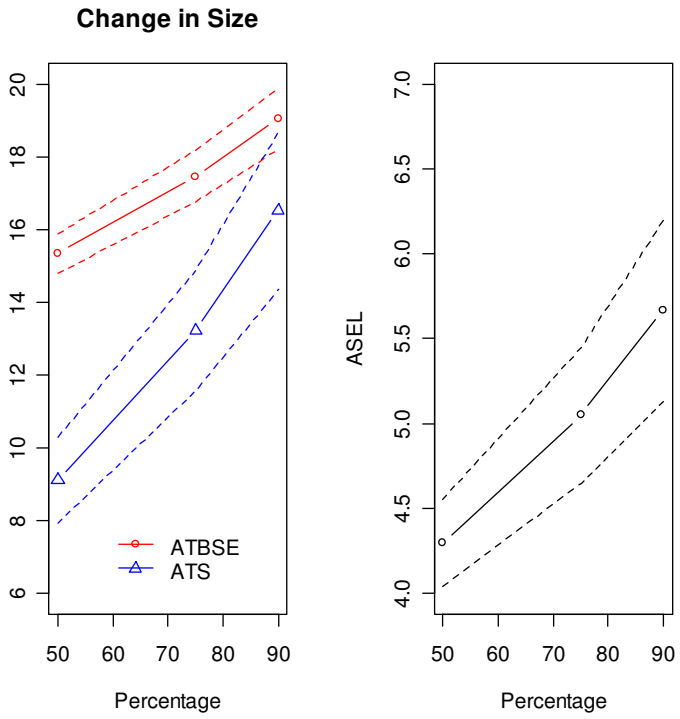


Figure 4.10: Effect of a change in the size limit of the sub-interval, %, on the in-control time-to-signal properties of the Kulldorff scan method. The rate is assumed to be $\mu = 10$.

The recurrence interval was also calculated. The value of the recurrence interval was approximately 5 for all combinations of μ, percentage, and S. This seems reasonable since $\alpha = 0.2$ corresponds to the marginal probability of a signal for each time period after the first. The recurrence interval is then roughly $1/\alpha = 5$. For this reason, it again seems that the recurrence interval is not summarizing the performance of the monitoring procedure in the same way as the time-to-signal measures.

It should be noted that the in-control ATBSE is always larger here than the in-control ATS meaning that once a signal has occurred the average time until another signal event is less than it was initially. This is not the same result that occurred when the properties of the non-

resetting EWMA chart were examined in Chapter 3. It is suspected that the reason for this somewhat counterintuitive result is the fact that the incidence rate is being estimated from the observed data. Once the incidence rate increases, the estimated incidence rate will also begin increasing as more time periods are collected at the higher rate. For data where there is no increase in the incidence rate, this could occur if several consecutive observations are above the estimated rate from the previous time periods. This is similar to what was occurring in Figure 4.4. This suggests that the method will discontinue signaling once the estimated incidence rate increases after enough time periods have occurred at the higher incidence rate.

4.3: Evaluation Under an Increase in the Incidence Rate

Kulldorff's scan method was also evaluated under a sustained increase in the incidence rate and an increase in the incidence rate of limited duration. The sustained increase in the incidence rate was assumed to start after 10, 50, or 200 time periods of the 250 total time periods when the initial incidence rate was 5. It started after 10, 50, or 100 time period of the 250 total time period when the initial incidence rate was 50. The limited shift began after 10 or 100 time periods and lasted for either 5 or 10 time periods. For an initial rate of 5 incidences per time interval, an increase to 7, 10 or 15 incidences per time period was examined. For an initial rate of 50 incidences per time interval, an increase to 75 or 100 incidences per time period was considered. For all combinations, 1,000 such datasets were generated. Since it was determined that the number of MC replications did not make a significant difference in the time-to-signal measures, $S = 999$ was used in the interest of saving computational time.

4.3.1: Sustained Increase

When there is a sustained increase in the incidence rate, the emphasis should be placed on quick detection. The sooner the increased incidence rate can be identified, the sooner corrective action or notification of individuals can take place. If a signal is produced it is hoped that it is a true indication of an increase in the incidence rate and not a false alarm since it can be costly to explore the causes of signals. Therefore, when summarizing the simulations of Kulldorff's scan method with a sustained increase in the incidence rate, the number of false alarms and the number of false alarm events were calculated. The number of false alarms is the number of time periods that result in a signal prior to the change in the incidence rate. The number of false alarm events is the number of signal events prior to the change in the incidence rate. Both of these metrics include iterations that do not contain any false alarms. Also calculated was the conditional expected delay (CED), defined as the expected time between the change in the incidence rate and the time period a signal is produced. The number of iterations that went into the calculation of the CED is also included since iterations with false alarms are excluded from this calculation. Table 4.2 shows the results of these calculations for both the adjusted and unadjusted cases when $\alpha = 0.2$ and when the maximum size of the sub-intervals being considered as a possible cluster is no greater than 50% of the total time periods. An X in the table indicates that the value can not be calculated.

Since the adjusted *p*-values are always at least as large as the unadjusted *p*-values, the number of false alarms and false alarm events is smaller for the adjusted case. The number of false alarm events is always no larger than the number of false alarms. A majority of the charts do not contain any false alarms or false alarm events in the adjusted case but some of these charts

do not result in a signal at all in the 250 time periods depending on the initial incidence rate and the amount of increase in the incidence rate. In the unadjusted case, the majority of the charts do result in at least one false alarm event. For the later change points, it is impossible to calculated the CED in the unadjusted case when the initial incidence rate is 5 because all of the charts contained at least one false alarm event. The CED for the adjusted case is always going to be at least as large as the unadjusted case. The hope in either case, though, is that the CED is small. The smaller the CED, the better the method is doing in detecting the increased incidence rate in terms of timely detection.

Intuitively, larger increases in the incidence rate, those of say an increase in the incidence rate of 5 to 15 or 50 to 100, should be detected more quickly then the smaller increases in the incidence rate, say an increase in the incidence rate from 5 to 7. When the initial incidence rate is 50, the increase in the incidence rate is detected immediately when using the unadjusted method except for the early increase in the incidence rate from 50 to 75. The adjusted *p*-value method has a difficult time detecting the small increase in the incidence rate from 5 to 7.

The time at which the incidence rate increases also appears to make a difference in the ability of the method to detect an increase in the incidence rate using the adjusted method. The later the increase occurs, the longer the scan method takes to detect the change. There does not seem to be a significant change in the CED in the unadjusted case as the time of the shift changes.

Based on the evaluation of the Kulldorff scan method in the in-control setting, the only factor that made a difference in the in-control time-to-signal metrics was the maximum size of the possible clusters. Therefore, the change in the incidence rate from 5 to 10 was also examined with an assumed percentage of 90%. The results can be found in Table 4.3 along with the results from a change in the incidence rate from 5 to 10 with an assumed percentage of 50%.

The conclusions of the comparison between the adjusted and unadjusted *p*-value methods for the maximum size of a cluster of 90% of the total number of time periods is the same as those from Table 4.2 where the percentage was set to be 50% of the total number of time periods. The mean number of false alarms is greater when clusters are allowed to be larger, but the mean number of false alarm events is smaller. This means that more of the false alarms occur at consecutive time intervals. This would result in fewer false inquiries into possible causes of an increased incidence rate by those monitoring the system. The CED is slightly larger for the larger percentage.

			Adjusted *p*-value Approach				Unadjusted *p*-value Approach			
μ_1	μ_2	*Change point*	*No. False Alarms*	*No. False Alarm Events*	*CED*	*No. with no False Alarms*	*No. False Alarms*	*No. False Alarm Events*	*CED*	*No. with no False Alarms*
5	7	11	0.494 (0.0294)	0.349 (0.0183)	90 do not signal	701	1.843 (0.055)	1.003 (0.0244)	3.754 (0.442)	269
5	7	51	1.206 (0.0729)	0.627 (0.0265)	2 do not signal	549	9.942 (0.230)	3.368 (0.0500)	3.875 (1.091)	16
5	7	201	2.243 (0.1364)	1.025 (0.0377)	5 do not signal	409	40.35 (0.823)	10.080 (0.1018)	X	0
5	10	11	0.476 (0.029)	0.318 (0.0165)	1.718 (0.114)	710	1.703 (0.053)	0.948 (0.0241)	0.539 (0.055)	293
5	10	51	1.296 (0.076)	0.670 (0.0275)	1.952 (0.079)	524	10.196 (0.235)	3.341 (0.0503)	0.421 (0.116)	19
5	10	201	2.305 (0.170)	1.023 (0.0399)	2.411 (0.087)	425	39.94 (0.798)	9.920 (0.0993)	X	0
5	15	11	0.464 (0.028)	0.311 (0.0163)	0.181 (0.016)	714	1.766 (0.056)	0.949 (0.0249)	0.026 (0.009)	306
5	15	51	1.241 (0.074)	0.651 (0.0274)	0.316 (0.022)	545	9.649 (0.230)	3.311 (0.0514)	0.095 (0.066)	21
5	15	201	2.526 (0.182)	1.074 (0.0402)	0.494 (0.032)	403	40.248 (0.782)	10.048 (0.1013)	X	0
50	75	11	0.524 (0.028)	0.390 (0.0189)	0.262 (0.021)	669	1.885 (0.056)	1.004 (0.0243)	0.051 (0.015)	254
50	75	51	1.191 (0.069)	0.673 (0.0267)	0.415 (0.025)	527	9.734 (0.230)	3.464 (0.0535)	0 (0)	25
50	75	101	1.707 (0.105)	0.854 (0.0348)	0.598 (0.032)	465	19.588 (0.437)	5.793 (0.0736)	0 (0)	2
50	100	11	0.49 (0.028)	0.352 (0.0177)	0.001 (0.001)	688	1.794 (0.054)	0.994 (0.0254)	0 (0)	289
50	100	51	1.318 (0.077)	0.705 (0.0284)	0.002 (0.002)	512	9.565 (0.221)	3.406 (0.0498)	0 (0)	17
50	100	101	1.737 (0.107)	0.883 (0.0344)	0.007 (0.004)	449	19.901 (0.449)	5.889 (0.0733)	0 (0)	2

Table 4.2: Summary of the sustained increase in the incidence rate. Here percentage = 50%, S = 999 and $\alpha = 0.2$. Standard error values are given in parentheses.

		Adjusted *p*-value Approach				Unadjusted *p*-value Approach			
Percentage	*Change Point*	*No. False Alarms*	*No. False Alarm Events*	*CED*	*No. with no False Alarms*	*No. False Alarms*	*No. False Alarm Events*	*CED*	*No. with no False Alarms*
50%	11	0.476 (0.029)	0.318 (0.0165)	1.718 (0.114)	710	1.703 (0.053)	0.948 (0.0241)	0.539 (0.055)	293
	51	1.296 (0.076)	0.670 (0.0275)	1.952 (0.079)	524	10.196 (0.235)	3.341 (0.0503)	0.421 (0.116)	19
	201	2.305 (0.170)	1.023 (0.0399)	2.411 (0.087)	425	39.94 (0.798)	9.920 (0.0993)	X	0
90%	11	0.613 (0.0462)	0.285 (0.0162)	1.868 (0.1057)	743	1.752 (0.0757)	0.675 (0.0217)	0.743 (0.0605)	443
	51	1.884 (0.1509)	0.669 (0.0312)	1.971 (0.0764)	569	10.203 (0.3584)	2.392 (0.0465)	0.700 (0.1172)	60
	201	3.001 (0.2770)	0.986 (0.0418)	2.4784 (0.0825)	464	40.65 (1.2537)	7.457 (0.1042)	0.250 (0.25)	4

Table 4.3: Summary of the performance of the sustained increase in the incidence rate. Here the change in the incidence rate is from 5 to 10 with S = 999 and $\alpha = 0.2$. Standard error values are given in parentheses.

4.3.2: Increase of Limited Duration

The Kulldorff scan method is also examined under an increase in the incidence rate of limited duration. With this type of increase in the incidence rate, the goal is to detect the increase quickly and while the incidence rate is still abnormally high. The method should also return to a non-signal state once the incidence rate has returned to normal. Table 4.4 shows the number of false alarms, number of false alarm events, CED, the number of iterations included in the calculation of the CED and when the time at which the process returns to a non-signal state after the incidence rate returns to the original value. Again, an X indicates the value could not be calculated.

As with the sustained increase in the incidence rate, the scan method is quicker in detecting a larger increase in the incidence rate. It is slower, however, in returning to an non-signal state when the shift is larger. This is mostly likely due to the fact that the incidence rate is being estimated from the data. When the increase in the incidence rate is larger, the estimated incidence rate will be larger and will require more time to return to the original incidence rate.

The time of the shift has an effect on the ability of the Kulldorff scan method to detect the change in the incidence rate and return to a non-signal state. Increases in the incidence rate that occur later in the monitoring process take longer to detect and take longer than those at earlier time periods to return to a non-signal state. The duration of the shift also has an effect of the metrics. The longer the increase in the incidence rate lasts, the longer the method takes to return to a non-signal state.

				Adjusted *p*-value Approach					Unadjusted *p*-value Approach				
μ_1	μ_2	*Change Point 1*	*Change Point 2*	*No. False Alarms*	*No. False Alarm Events*	*CED*	*No. with no False Alarms*	*Returns to Non-Signal State*	*No. False Alarms*	*No. False Alarm Events*	*CED*	*No. with no False Alarms*	*Returns to Non-Signal State*
5	10	11	16	0.501 (0.0293)	0.339 (0.0171)	X	693	19.024 (0.0742)	1.816 (0.0538)	0.991 (0.0240)	1.0632 (0.4060)	269	21.450 (0.0741)
5	15	11	16	0.477 (0.0273)	0.350 (0.0178)	0.175 (0.0161)	692	21.807 (0.0396)	1.878 (0.0565)	0.993 (0.0251)	0.0283 (0.0099)	283	23.043 (0.0398)
5	10	101	111	1.803 (0.1136)	0.858 (0.0323)	2.1991 (0.0792)	447	139.158 (0.6579)	20.374 (0.4572)	5.770 (0.0721)	1.667 (0.8819)	3	X
5	15	101	111	1.777 (0.1200)	0.843 (0.0324)	0.447 (0.0276)	470	183.218 (0.5976)	20.128 (0.4595)	5.626 (0.0724)	0 (0)	3	201.319 (0.2867)
5	10	11	21	0.439 (0.0256)	0.330 (0.0173)	X	707	23.714 (0.0625)	1.773 (0.0542)	0.997 (0.0254)	0.474 (0.0483)	287	26.085 (0.0735)
5	15	11	21	0.481 (0.0287)	0.346 (0.0180)	0.214 (0.0184)	698	26.171 (0.0455)	1.892 (0.0582)	1.011 (0.0256)	0.043 (0.0121)	280	27.714 (0.0480)

Table 4.4: Summary of the limited duration increase in the incidence rate. Here the percentage = 50%, S = 999 and $\alpha = 0.2$. Standard error values are given in parentheses.

4.4: Discussion

Several different factors were examined in terms of their effect on the Kulldorff scan method. Neither the initial incidence rate nor the number of MC simulations performed affected the in-control time-to-signal properties or the value of the recurrence interval. The maximum size of the possible clusters, however, did significantly affect the in-control time-to-signal metrics, but not the recurrence interval. Larger values of the percentage resulted in larger in-control time-to-signal values.

The use of the adjusted method seems to be inadequate in determining if there are possible times of higher than normal incidence rate. The adjusted method results in *p*-values that are close to one and, therefore, a signal is almost never obtained. This sometimes results in a long delay in detecting an increase in the incidence while at other times, the increase in the incidence rate is never signaled. The unadjusted method seems to detect changes in the incidence rate in a more timely manner without a large number of false alarm events.

In the cases where an increase in the incidence rate was known to have occurred, allowing larger clusters resulted in a slightly longer CED, but not significantly so. The time at which the change in the incidence rate occurs and the size of the increase in the incidence rate were also important factors for both type of increases in the incidence rate. The larger the increase in the incidence rate, the quicker the method detects the increase, as would be expected.

Chapter 5

Comparison of Kulldorff's Scan Method with Other Surveillance Schemes

There have been several methods described in the industrial SPC literature to monitor attribute data (see Woodall, 1997). Many of these methods are based on the assumption that the counts follow a Poisson distribution. Some of these methods will be discussed in this chapter. The performance of the scan method proposed by Kulldorff (1997 and 2001) will be compared to the performance of these methods.

5.1: The c-chart

The c-chart is commonly used to monitor the number of defects in a manufacturing process. It could be applied to public health surveillance data by monitoring the number of incidences of a disease or other variable of interest. The details of the c-chart can be found in Montgomery (2005). The assumptions for the use of this chart are that the data follow a Poisson

distribution and each sub-group has an equal size. In terms of the temporal disease rate monitoring, this means that the aggregated time units are of equal size.

For the c-chart, the chart centerline is the mean number of incidences, $\bar{c}$, in the Phase I sample. The control limits are given by

$$LCL = \max\left(0, \bar{c} - 3\sqrt{\bar{c}}\right) \tag{5.1}$$

$$UCL = \bar{c} + 3\sqrt{\bar{c}} \tag{5.2}$$

A signal in the change of the Poisson rate is produced if the count falls outside of the control limits. If, however, the rate of nonconformities is known, then the true rate, *c*, is used instead of the estimated rate from the Phase I observations.

5.2: The Poisson CUSUM Chart

The details for the use of the Poisson CUSUM chart were given by Lucas (1985). As with the c-chart, the data collected are counts of incidences assumed to be Poisson distributed. The control chart statistic for detecting an increase in the incidence rate is

$$S_i = \max(0, Y_i - k + S_{i-1}) \tag{5.5}$$

for $i = 0, 1, 2, \ldots$ and $S_0 = 0$ for the standard Poisson CUSUM chart or $S_0 = h/2$ for a fast initial response (FIR) CUSUM chart. The reference value, *k*, is determined using the formula

$$k = (\mu_d - \mu_a) / (\ln(\mu_d) - \ln(\mu_a)) \quad (5.6)$$

where μ_a is the in-control process mean, μ_d is the mean that the CUSUM is to quickly detect, and ln(.) is the natural log. The value of the cut-off *h* that gives a desired in-control ATS can then be found using simulation. For the comparisons performed in the next section, *h* is chosen so that the in-control ATS for the Poisson CUSUM chart is the same as the in-control ATS for Kulldorff's method with the same value of μ_a. A signal is produced if the value of the Poisson CUSUM statistic meets or exceeds *h*.

5.3: Kulldorff's Scan Method vs. the Poisson CUSUM Chart

White et al. (1997) showed that the Poisson CUSUM chart is superior to the c-chart in terms of average run length performance. In this section, therefore, the statistical performance of the standard Poisson CUSUM chart will be compared to that of Kulldorff's scan method.

5.3.1: Sustained Increase

In the sustained increase in the incidence rate situation, the same data that were used for the evaluation of the Kulldorff scan method are examined using the Poisson CUSUM chart. For an initial incidence rate of 5, *h* was set to be 2.1. This resulted in an in-control ATS for the Poisson CUSUM chart of 8.832 with a standard error of 0.2494, compared to the in-control ATS for the Kulldorff scan method with an incidence rate of 5 and percentage equal to 50% which was found to be 8.935 with a standard error of 0.4091. For an incidence rate of 10, *h* was found to be between 0.78 and 0.79 which resulted in an in-control ATS of 7.404 with a standard error of 0.2233 and 14.415 with a standard error of 0.4644, respectively. The in-control ATS for the

Kulldorff scan method with an incidence rate of 10 with a percentage of 50% was 9.126 with a standard error of 0.3913.

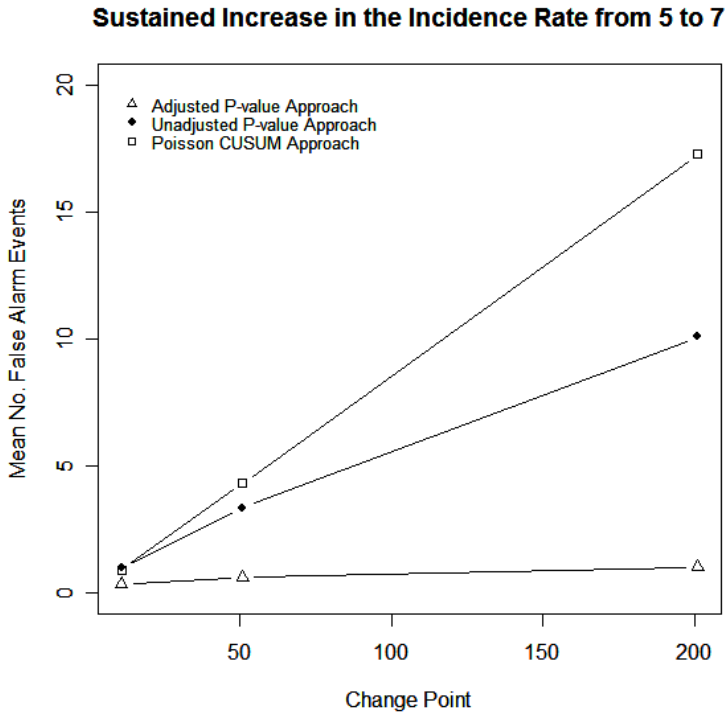
Table 5.1 shows the same metrics as Table 4.2. Figure 5.1 shows the mean number of false alarm events for both the adjusted and unadjusted *p*-value approach to the Kulldorff scan method and the Poisson CUSUM chart for the different change points for (a) an increase in the incidence rate from 5 to 7 and (b) for an increase in the incidence rate from 5 to 10. Figure 5.2 shows the CED for both the adjusted and unadjusted *p*-value approach to the Kulldorff scan method and the Poisson CUSUM chart for the different change points for (a) an increase in the incidence rate from 5 to 7 and (b) for an increase in the incidence rate from 5 to 10. Comparing the two Kulldorff scan methods, the adjusted *p*-value approach and the unadjusted *p*-value approach, and the Poisson CUSUM chart for the increase in the incidence rate from 5 to 7 shows that the Poisson CUSUM chart has more false alarm events than the scan methods for later shifts in the incidence rate. This is due to the small in-control ATS of the Poisson CUSUM chart. The Poisson CUSUM chart, however, detects the change in the incidence rate faster than the Kulldorff scan method. It should be noted that the CED comparison between the Poisson CUSUM chart and Kulldorff's method in this case is done only for the unadjusted *p*-value method. This information was not available for the adjusted *p*-value method since some scan method iterations never produced signals.

The mean number of false alarm events and the CED for an increase in the incidence rate from 5 to 10 can be compared to both the adjusted and unadjusted *p*-value approach. Again, the Poisson CUSUM chart is quicker in detecting the increase in the incidence rate using either value

of h, though the unadjusted p-value approach of Kulldorff is only slightly slower. The adjusted p-value approach of the Kulldorff method has the fewest number of false alarm events and the unadjusted p-value approach of the Kulldorff method the most false alarm events for earlier increases in the incidence rate. For the latest increase in the incidence rate, the Poisson CUSUM chart has the largest number of false alarm events; again due to the small in-control ATS.

μ_1	μ_2	*Change Point*	h	*No. False Alarms*	*No. False Alarm Events*	*CED*	*No. with no False Alarms*
5	7	11	2.1	2.149 (0.0700)	0.925 (0.0243)	1.497 (0.1063)	312
5	7	51	2.1	13.751 (0.2215)	4.324 (0.0508)	0.5 (0.5000)	2
5	7	201	2.1	57.630 (0.4725)	17.260 (0.0996)	X	0
5	10	11	0.79	1.092 (0.0435)	0.692 (0.0231)	0.4811 (0.0391)	449
5	10	51	0.79	5.527 (0.1000)	3.329 (0.0506)	0.4 (0.1338)	20
5	10	201	0.79	22.371 (0.2159)	13.176 (0.0996)	X	0
5	10	11	0.78	1.644 (0.0467)	1.175 (0.0280)	0.3034 (0.0198)	234
5	10	51	0.78	8.230 (0.1067)	5.657 (0.0596)	X	0
5	10	201	0.78	32.971 (0.2229)	22.371 (0.1153)	X	0

Table 5.1: Summary of the performance of the Poisson CUSUM chart with a sustained increase in the incidence rate. Standard errors are given in the parentheses.



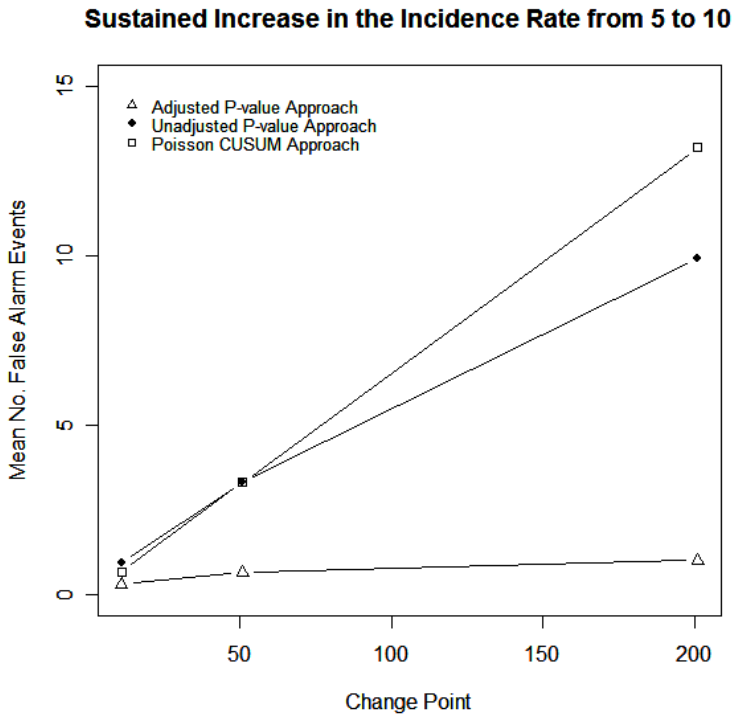
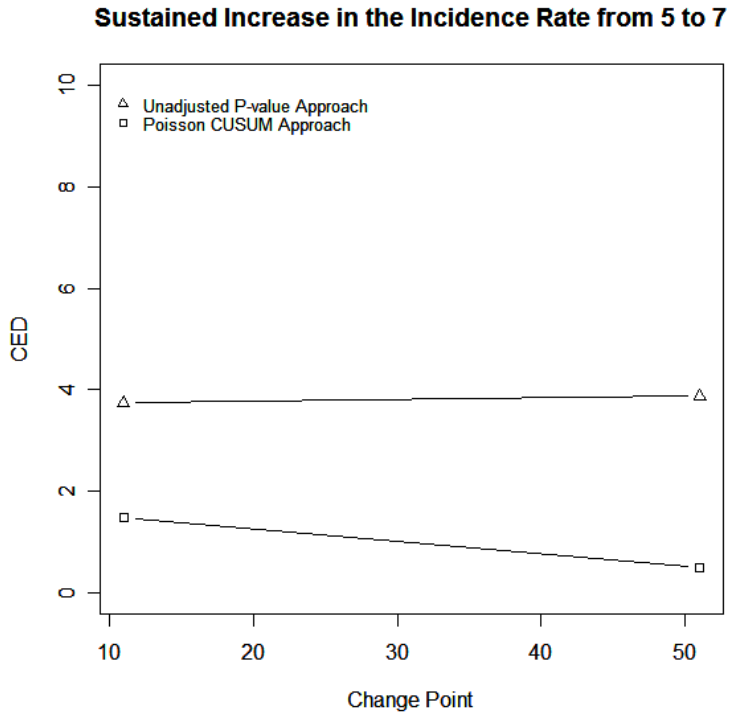


Figure 5.1: Change in the mean number of false alarm events as the time at which the increase in the incidence rate increase from (a) a rate of 5 to 7 and (b) a rate of 5 to 10.



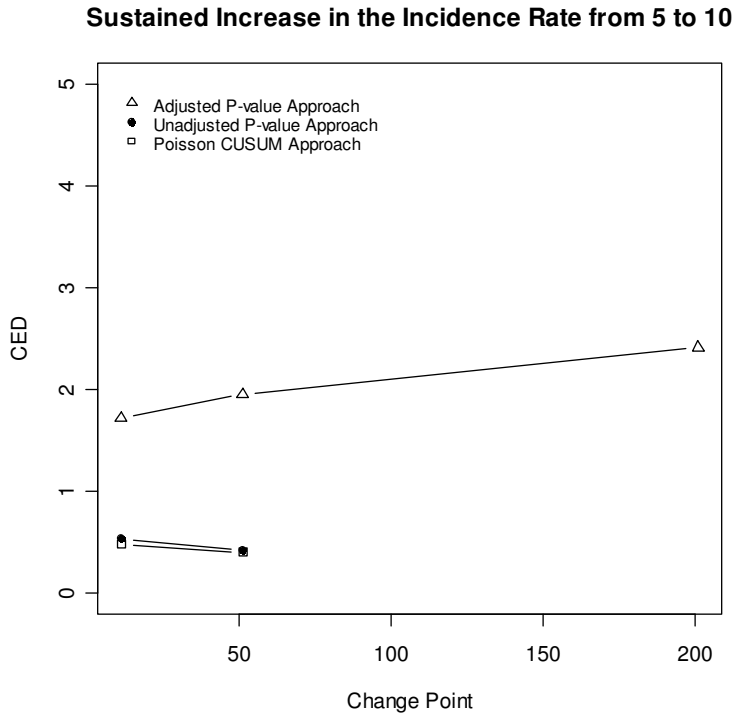


Figure 5.2: Change in the CED as the time at which the increase in the incidence rate increase from (a) a rate of 5 to 7 and (b) a rate of 5 to 10.

5.3.2: Increase of Limited Duration

In the situation of an increase in the incidence rate of limited duration, the same data that were used for the evaluation of the Kulldorff scan method was examined using the Poisson CUSUM chart. The same cut-off values are used for the Poisson CUSUM chart as were used in the examination of the properties under a sustained increase in the incidence rate. The results in this case are roughly the same as those from the sustained increase in the incidence rate and are given in Table 5.2. The Poisson CUSUM chart detects the change in the incidence rate faster, on average, than either the adjusted or unadjusted *p*-value Kulldorff method. The Poisson CUSUM chart falls in between the adjusted and unadjusted *p*-value Kulldorff methods in terms of the average number of false alarm events for earlier increases in the incidence rate. The Poisson CUSUM chart returns to a non-signal state faster than the Kulldorff method after the incidence rate has returned to the original level.

μ_1	μ_2	*Change Point 1*	*Change Point 2*	*h*	*No. False Alarms*	*No. False Alarm Events*	*CED*	*No. with no False Alarms*	*Returns to Non-Signal State*
5	10	11	16	0.79	1.060 (0.0427)	0.689 (0.0229)	0.457 (0.0358)	451	16.546 (0.1044)
5	10	101	111	0.79	11.129 (0.1447)	6.636 (0.0712)	X	0	111.033 (0.0150)
5	10	11	21	0.79	1.007 (0.0393)	0.674 (0.0221)	0.4089 (0.0352)	450	21.001 (0.0010)
5	10	11	16	0.78	1.649 (0.0464)	1.205 (0.0283)	0.215 (0.0347)	228	16.332 (0.0720)
5	10	101	111	0.78	16.429 (0.1534)	11.242 (0.0842)	X	0	111.021 (0.1126)
5	10	11	21	0.78	1.581 (0.0422)	1.192 (0.0268)	0.268 (0.0181)	220	21.001 (0.0010)

Table 5.2: Summary of the performance of the Poisson CUSUM chart with an increase in the incidence rate of limited duration. Standard errors are given in parentheses.

5.4: Discussion

The Poisson CUSUM method seems to detect an increase in the incidence rate faster than the Kulldorff scan method for both a sustained increase in the incidence rate and for an increase in the incidence rate of limited duration. There is a trade off, however. The Poisson CUSUM chart sometimes results in more false alarm events. Part of the reason for the Poisson CUSUM chart detecting the increases in the incidence rate faster is the fact that the known baseline incidence rate is used with the Poisson CUSUM chart while an estimate of the incidence rate from the observed data is needed for the Kulldroff scan method. Using this information when it is available improves the time-to-signal performance of surveillance methods.

Chapter 6

Future Research Directions

Public health surveillance is a growing field of research interest. The evaluation of the Kulldorff (2001) methods presented in Chapters 4 and 5 is only a subset of the areas that need to be examined. Some suggestions to correct for various issues with the basic Kulldorff scan method are presented in this section while other areas are research are presented in alter sections.

6.1: Possible Improvements to the Kulldorff Scan Method

Based on the evaluations of the methods performed in Chapters 4 and 5, there appear to be some limitations with the scan method proposed by Kulldorff (2001). Some of these are based on the fact that the baseline incidence rate is estimated from the data. This is an artifact of conditioning on the total number of observed incidences. Some suggestions have been made to

correct for this. While these have not been examined here, they could be useful comparisons for the future.

6.1.1: Improving Power

In Kulldorff (2001) a suggestion is made that the author believed would improve the power of the proposed scan method. To do this, the amount of time that is kept in the "history" of the method is restricted to only a specific number of time units. For example, it is possible in a monitoring process to only consider the last 20 weeks. This is different than restricting the size of the possible clusters with the percentage of time periods allowed to be in a cluster. The limited history restricts the amount of past information used when calculating both the maximum observed LLR and the distribution of the maximum LLR estimated from the MC simulation. For instance, if there are 50 time periods of data available and the history is set to be the last 10 time periods, then the maximum size of the cluster when the percentage is set at 50% is 5 but these sub-intervals must appear in the last 10 time periods only. Sub-intervals occurring in the first 40 time periods are no longer considered. Therefore, any information prior to the last 10 time periods is lost. This is also the case in the estimation of the baseline incidence rate from the data.

6.1.2: Improving Computational Requirements

Sonesson (2007) showed that the Kulldorff scan method can be written in terms of a CUSUM framework where conditioning on N allows for the estimation of the incidence rate from the observed data by maximum likelihood estimation assuming there is no increase in the incidence rate. The justification from Kulldorff for the need to condition on the total number of observations, N, is to be able to obtain the p-values from the MC simulation. It is these MC

simulations, however, which require the majority of the computational time necessary to perform the Kulldorff scan method. With CUSUM charts, however, a cut-off value on the CUSUM statistic is determined and a signal is produced if the value of the CUSUM statistic is at least as large at the cut-off value. With this reasoning, it might be possible to determine a cut-off value for the maximum observed LLR surveillance statistic for the Kulldorff scan method and not perform the additional steps of calculating the *p*-values.

Figure 6.1 shows the observed number of incidences, adjusted *p*-values, unadjusted *p*-values, and the maximum observed LLR plotted over time for one iteration when $\mu = 10$, the percentage is 75% and S = 999. The unadjusted *p*-values and the maximum observed LLR plots appear to be mirror images of each other to some extent. That is, when the unadjusted *p*-value increases, the maximum observed LLR decreases and vice versa. In fact, the mean Pearson's correlation coefficient between these two values over all 1,000 iterations is -0.9031. A negative correlation is reasonable since it is a large maximum observed LLR that results in a small *p*-value when trying to detect increases in the incidence rate. Since these two values are so highly related, it seems reasonable that a cut-off for the maximum observed LLR value could be found instead of estimating the *p*-values based on the extensive amount of MC simulation.

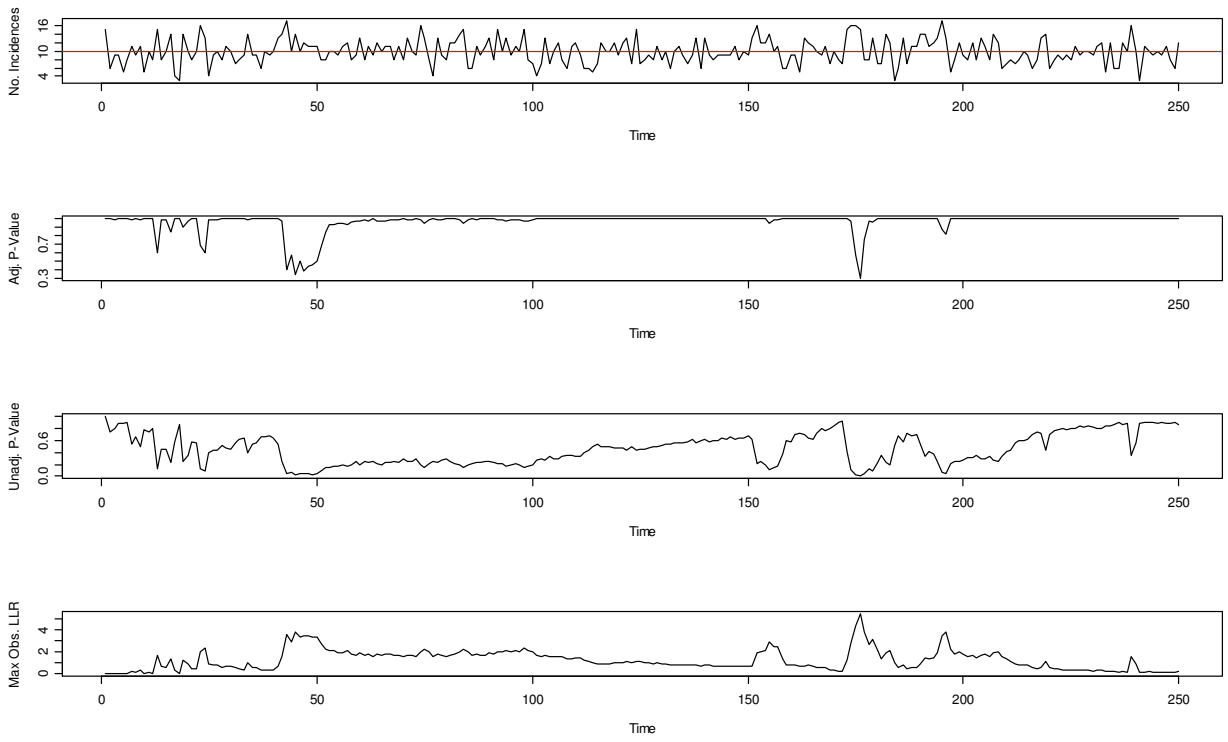


Figure 6.1: Plots of the observed number of incidences, adjusted and unadjusted *p*-values, and the maximum observed LLR surveillance statistic for one iteration of the simulation with $\mu = 10$, percent = 75%, and S = 999.

6.1.3: Unconditional Log-Likelihood Ratio

The Kulldorff scan method is based on the conditional log-likelihood ratio. Estimating the incidence rate from the data seems to have some disadvantages. If the incidence rate is known or can be estimated from previous studies, then information is lost when estimating the incidence rate from the observed data. There are other methods, that do not condition on the total number of observed incidences. Implementing these methods, the in-control incidence rate is assumed to be known. Sonesson (2007) made a comparison between some of these methods in the space-time setting. Here, the scan method of Kulldorff was compared to two different CUSUM methods, the nearest-neighborhood CUSUM of Raubertas (1989) and the circular

CUSUM. These differ in the way a possible cluster is defined but both have the property that the initial incidence rate is known.

Sonesson (2007) did caution that "if one has access to long, stable historical data, the unconditional approach will work well, since the estimation of λ_0 [the initial incidence rate] will be precise. However, some caution must be raised if the baseline risk is not well known." Often there is more background information available in the industrial applications than in the health care applications. Also, health care monitoring often suffers from underreporting. This can occur because reporting methods are not easy or simply because practitioners do not take the time to report all symptoms or problems. If it is known that underreporting is occurring, then using some known incidence rate, maybe from clinical trials or other earlier studies, could result in issues when monitoring the data with a method that assumes a known rate. Estimating the incidence rate from the data might result in slightly slower detection time, but it could also possibly help with the underreporting problem. Even though the estimated incidence rate might not be near the true incidence rate, when an increase does occur the Kulldorff method would be able to detect it when another method that was based on a known incidence rate might not.

6.2: Other Research Areas

6.2.1: The Binomial Case

The Poisson model is not the only distribution that the data can be assumed to have when using the Kulldorff scan method. The binomial case can also be used. The likelihood ratio for this situation is given in Kulldorff (1997). The properties of the prospective Kulldorff scan method under the assumption of a binomial process have not been studied.

6.2.2: The Marginal Probabilities of a Signal for the One-Sided CUSUM Chart

In Chapter 3 Section 2, it was shown by simulation that for $h = 4.0$ or 5.0 and $k = 0.5$ or 1.0, the marginal probabilities of a signal approach a limiting value. This fact allowed the recurrence interval to be defined in terms of steady state behavior for use with the one-sided CUSUM chart. The theoretical approach to show the marginal probabilities of a signal will always reach steady state would involve the use of a Markov process representation of the CUSUM chart. A similar approach was taken by Fu et al. (2003), among others, to examine the run-length distribution of the CUSUM chart. Since the CUSUM chart involves a continuous state space, it does not lend itself naturally to a Markov chain representation except as an approximation. The Poisson CUSUM chart, however, may have an exact Markov chain representation. The approximation is performed by creating equally spaced divisions between the origin and the signal limit. Typically 500 such divisions are assumed to be enough to obtain reasonable estimates. Any CUSUM statistic that falls beyond the signal limit falls into an absorbing state.

The approximation, however, would not work to calculate the marginal probability of a signal. The difference being that any CUSUM statistic that falls above the signal limit should not be considered as one state of the Markov Chain. Therefore, there will still be an infinite number of states.

References

Bhat, U. N. and Lal, R. (1990). Attribute Control Charts for Markov Dependent Production Processes. *IIE Transactions* 22:181-188.

Buehler, J. W., Hopkins, R. S., Overhage, J. M., Sosin, D. M., and Tong, V. (2004). Framework for Evaluating Public Health Surveillance Systems for Early Detection of Outbreaks, *Morbidity and Mortality Weekly Report* 53: 1-11.

CDC (2001). Updated Guidelines for Evaluating Public Health Surveillance Systems: Recommendations for the Guidelines Working Group, *Morbidity and Mortality Weekly Report* 50: 1-35.

Daniel, J. B., Heisey-Grove, D., Gadam, P., Yih, W., Mandl, K., DeMaria, Jr., A., and Platt, R. (2005). Connecting Health Departments and Providers: Syndromic Surveillance's Last Mile. *Morbidity and Mortality Weekly Report* 54 (Suppl.): 147-150.

Fraker, S.E, Woodall, W.H., and Burkom, H.S. (2007) A Note on the Poisson Likelihood Ratio Test Statistic for Kulldorff's Scan Methods. To appear in *Communications in Statistics – Theory and Methods.*

Fraker, S.E., Woodall, W.H., and Mousavi, S. (2007) Performance Metrics for Surveillance Schemes. Submitted to *Quality Engineering (special issue on health care).*

Frisén, M. and Sonesson, C. (2005). Optimal Surveillance. Chapter 3 in *Spatial & Syndromic Surveillance* by A. B. Lawson and K. Kleinman, John Wiley & Sons, Ltd.: New York, NY, 31-52.

Fu, J. C., Shmueli, G., and Chang, Y. M. (2003). A Unified Markov Chain Approach for Computing the Run Length Distribution in Control Chart with Simple or Compound Rules. *Statistics and Probability Letters* 65: 457-466

Grigg, O. and Farewell, V. (2004). An Overview of Risk-adjusted Charts. *Journal of the Royal Statistical Society* A: 167: 523-539.

Grigg, O. and Spiegelhalter, D. (2006). The Null Steady-State Distribution of the CUSUM Statistic. Submitted for Publication.

Joner, M. D., Jr., Woodall, W. H., and Reynolds, M. R., Jr. (2007). On Detecting a Rate Increase Using a Bernoulli-based Scan Statistic. To appear in *Statistics in Medicine.*

Kleinman, K. (2005). Generalized Linear Models and Generalized Linear Mixed Models for Small-Area Surveillance. Chapter 5 in *Spatial & Syndromic Surveillance* by A. B. Lawson and K. Kleinman, John Wiley & Sons, Ltd.: New York, NY, 77-94.

Kleinman, K. P., Abrams, A. M., Kulldorff, M., and Platt, R. (2005). A Model-Adjusted Space-Time Scan Statistic with an Application to Syndromic Surveillance. *Epidemiology and Infectious Diseases* 133: 409-419.

Kleinman, K. P., Abrams, A., Mandl, K., and Platt, R. (2005). Simulation for Assessing Statistical Methods of Biologic Terrorism Surveillance. *Morbidity and Mortality Weekly Report* 54 (Suppl): 101-108.

Kleinman, K., Lazarus, R., and Platt, R. (2004). A Generalized Linear Models Approach for Detecting Incident Clusters of Disease in Small Areas, with an Application to Biological Terrorism (with discussion). *American Journal of Epidemiology* 159: 217-228.

Kulldorff, M. (2005a). SaTScan: Software for the spatial, temporal, and space-time scan statistics, version 5.1 [computer program]. Information Management Services 2005; Available: http://www.satscan.org/.

Kulldorff, M. (2005b). Scan Statistics for Geographical Disease Surveillance: An Overview. Chapter 7 in *Spatial & Syndromic Surveillance* by A. B. Lawson and K. Kleinman, John Wiley & Sons, Ltd.: New York, NY, 115-132.

Kulldorff, M. (2001). Prospective Time Periodic Geographical Disease Surveillance Using a Scan Statistic. *Journal of the Royal Statistical Society*, Series A, 164: 61-72.

Kulldorff, M. (1997). A Spatial Scan Statistic. *Communications in Statistics – Theory and Methods*, 26: 1481-1496.

Kulldorff, M., Tango, T., and Park, P. J. (2003) Power Comparisons for Disease Clustering Tests. *Computational Statistics and Data Analysis* 42: 665-684.

Kulldorff, M., Zhang, Z., Hartman, J., Heffernan, R., Huang, L., and Mostashari, F. (2004) Benchmark data and power calculations for evaluating disease outbreak detection methods. *Morbidity and Mortality Weekly Report*, **53** (Supplement), 144-151.

Lucas, J. M. (1985). Counted Data CUSUMs. *Technometrics* 27(2): 129-144

Margavio, T. M., Conerly, M. D., Woodall, W. H., and Drake, L. G. (1995). Alarm Rates for Quality Control Charts. *Statistics and Probability Letters* 24: 219-224.

Montgomery, D. C. (2005). *Introduction to Statistical Quality Control 5th Ed*. John Wiley & Sons, Ltd.: New York, NY.

Mousavi, Sabnam M. (2006). Monitoring Markov Dependent Binary Observations with a Log-Likelihood Base CUSUM Control Chart.

Naus J. and Wallenstein S. (2006). Temporal Surveillance Using Scan Statistics. *Statistics in Medicine*: 25: 311-324.

Nordin, J. D., Goodman, M. J., Kulldorff, M., Ritzwoller, D. P., Abrams, A. M., Kleinman, K., Levitt, M. J., Donahue, J., and Platt, R. (2005). Simulated Anthrax Attacks and Syndromic Surveillance. *Emerging Infectious Diseases* 11: 1394-1398.

Page, E. S. (1954). Continuous Inspection Schemes. *Biometrika* 41: 100-115.

Przyborowski, J. and Wilenski, H. (1940). Homogeneity of Results in Testing Sample from Poisson Series: With an Application to Testing Clover Seed for Dodder. *Biometrika* 31: 313-323.

Raubertas, R. F. (1989). An Analysis of Disease Surveillance Data That Uses the Geographic Locations of the Reporting Units. *Statistics in Medicine* 8(3): 267-271.

Reis, B. Y. and Mandl, K. D. (2003). Time Series Modeling for Syndromic Surveillance. *BMC Medical Informatics and Decision Making*: 3(2).

Reynolds, M. R., Jr. and Stoumbos, Z. G. (1999). A CUSUM Chart for Monitoring a Proportion When Inspecting Continuously. *Journal of Quality Technology* 31: 87-108.

Roberts, S. W. (1959). Control Chart Tests Based on Geometric Moving Averages. *Technometrics* 1(3): 239 – 250.

Shmueli, G. (2007). Statistical Challenges in Modern Biosurveillance. Invited paper submitted to *Technometrics*.

Song, C. and Kulldorff, M. (2003) Power Evaluation of Disease Clustering Tests. *International Journal of Health Geographics* 2: 1-8.

Steiner, S. H. (1999). EWMA Control Charts with Time-Varying Control Limits and Fast Initial Response. *Journal of Quality Technology* 31: 75-86.

Steiner, S. H., Cook, R. J., Farewell, V. T., Treasure, T. (2000). Monitoring Surgical Performance Using Risk-adjusted Cumulative Sum Charts. *Biostatistics* 1: 441-452.

Sonesson, C. (2007) A CUSUM Framework for Detection of Space-Time Disease Clusters Using Scan Statistics. To appear in *Statistics in Medicine.*

Stoto, M. A., Fricker, R. D., Jr., Jain, A., Diamond, A., Davies-Cole, J. O., Glymph, C., Kidane, G., Lum, G., Jones, L., Dehan, K., and Yuan, C. (2006). Evaluating Statistical Methods for Syndromic Surveillance, chapter in *Statistical Methods in Counterterrorism* edited by A. G. Wilson, G. D. Wilson, and D. H. Olwell, Springer: 141-172.

Waller, L. A. (2004). Invited Commentary: Surveilling Surveillance – Some Statistical Comments. *American Journal of Epidemiology* 159: 225-227.

Wallenstein, S. and Naus, J. (2004). Scan Statistics for Temporal Surveillance for Biologic Terrorism. *Morbidity and Mortality Weekly Report*: 53 (Suppl): 74-78.

White, C. H., Keats, J. B., and Stanley, J. (1997). Poisson CUSUM versus c Chart for Defect Data. *Quality Engineering* 9(4): 673-679.

Woodall, W. H., Marshall, J. B., Joner, M.D., Jr., Fraker, S.E., and Abdel-Salam, A.G. (2007) On the Use and Evaluation of Scan Methods in Prospective Public Health Surveillance. To appear in the *Journal of the Royal Statistical Society*, Series A.

Woodall, W. H. (2006). The Use of Control Charts in Health-Care and Public-Health Surveillance. *Journal of Quality Technology* 32: 88-103.

Woodall, W. H. (1997) Control charts based on attribute data: Bibliography and review. *Journal of Quality Technology* 29: 172 - 183.

Yih, W. K., Caldwell, B., Harmon, R., Kleinman, K., Lazarus, R., Nelson, A., Nordin, J., Rehm, B., Richter, B., Ritzwoller, D., Sherwood, E., and Platt, R. (2004). National Bioterrorism Syndromic Surveillance Demonstration Program. *Morbidity and Mortality Weekly Report* 53 (Suppl.): 43-49.

Zhu, Y., Wang, W., Atrubib, D., and Wu, T. (2005) Initial evaluation of the early aberration reporting system – Florida. *Morbidity and Mortality Weekly Report*, 54 (Suppl.): 123-130.