Properties and Use of the Shewhart Method and its Followers

Marianne Frisén

Statistical Research Unit, Department of Economics, Göteborg University
Göteborg, Sweden

Abstract: After the Shewhart method was suggested for industrial applications, other applications such as surveillance for bioterrorism and financial transactions have come into focus. Also other methods for surveillance have followed. The relation between the Shewhart method and the followers is examined. A uniform presentation of methods, by expressions of likelihood ratios, facilitates the comparisons between methods. The situations for which the Shewhart method has optimality properties are thus determined. The uses of the Shewhart method and its followers for complicated situations are reviewed.

Keywords: Change-point detection; Control chart; Likelihood ratio; Monitoring; Quality control; Shewhart; Statistical process control; Surveillance;

Subject Classifications: 62-02; 62C10; 62C20; 62L15.

Address correspondence to: Marianne Frisén, Statistical Research Unit, Göteborg University, Box 660, SE-40530 Göteborg, Sweden; Fax: 46-31-7731274; E-mail: marianne.frisen@statistics.gu.se

1 INTRODUCTION

In the 1920s, Walter A. Shewhart and co-workers at Bell Telephone laboratories developed the first versions of a modern control chart. In 1931 Shewhart published his famous book “Economic Control of Quality of Manufactured Product” (Shewhart (1931)). The same year he gave a presentation of the new technique to the Royal Statistical Society. This stimulated interest in the UK. The technique was used extensively during World War II both in the UK and in the US. In the 1950s, W. E. Deming introduced the technique in Japan. The success in Japan spurred the interest in the West, and further development started.

The need is great for continual observation of time series with the aim of detecting an important change in the underlying process as soon as possible after it has occurred. The statistical methods suitable for surveillance differ from the standard hypothesis testing methods. Also the criteria for optimality differ from those used for ordinary hypothesis testing. Evaluations and optimality criteria are important in order to choose which surveillance method to use for a specific purpose. For example, different requirements apply for short-term high-risk situations as compared to long-time low-risk ones.

Examples of broad surveys and bibliographies on statistical surveillance are found in the following papers: Zacks (1983) points out that surveillance is a central problem of statistical inference, linking together different areas of statistical theory. Yashchin (1993) discusses the relation between “Engineering Process Control” where the corrective formula is important and “Statistical Process Control”, SPC, where the detection of the abrupt change is the main aim. Woodall and Montgomery (1999) and Stoumbos et al. (2000) concentrate on SPC but stress that a cross-fertilization with the mathematical statistical literature on change-point analysis would be fruitful. Frisén (2003) gives a survey with concentration on the link between methods and optimality criteria. The two discussion papers Lai (2001) and Lai (2004) contain important contributions on recent and new challenges, especially on asymptotic optimalities of generalized likelihood ratio methods for surveillance.

Although industrial applications are still important, new applications have come into focus. Such new applications are described in Section 0. The Shewhart method, described in Shewhart (1931), is simple and certainly the most commonly used method for surveillance. In Section 3 the Shewhart method is described. Moreover, a description is made of the most commonly discussed case in the literature, that of a shift in the mean of a normal distribution. This case is used in Section 4 and 5 to illustrate some measures for evaluations and some optimality criteria. In Section 6 we describe the need for developments after the Shewhart method and the optimality of the followers. Most methods are optimal in some respect. One of the methods described is the full likelihood ratio method, LR. The LR method corresponds to the use of the posterior probability of a change and fulfils important optimality criteria. This method, then, can be used as a benchmark for the Shewhart method as well as other methods, since they all can be expressed by partial likelihoods. In Section 7 we demonstrate some relations between the Shewhart method and other methods for surveillance derived from different optimality criteria. We can thus present stochastic properties of the Shewhart method derived from its relation to other methods which have known optimality properties. In Section 8 the use of the Shewhart method and its followers is discussed with regard to some more complicated situations, like multivariate surveillance, more complicated models and more complicated changes. Section 9 contains some concluding remarks.

2 THE NEED FOR SURVEILLANCE IN NEW AREAS

Shewhart gave his suggestions motivated by the needs in industrial quality control. He was at that time a researcher with Bell Laboratories and recognized that no matter how precise a manufacturing process was, in the real world it could never produce exactly the same output each time. Thus statistical methods are necessary to judge how much information we have about a possible change from the target. Comments on the role of statistical quality control in industry are given in the paper by Banks (1993) and the connected discussion. In industry new applications such as management problems, described for example by Pettersson (2004), have been of recent interest. Moreover, Tartakovsky and Veeravalli (2002) described how surveillance can be used to detect targets in multichannel and multisensor distributed systems as well as to build systems for early detection of intrusions in computer networks.

Shewhart recognized the fact that his method is general and that its application is not restricted to manufacture. Applications in quite different areas have come into focus lately and have influenced the theoretical research on methods and evaluations. In recent years, there have been a growing number of papers in medicine, environmental control, economics and other areas dealing with the need for methods for surveillance.

A very recent and very important issue is the surveillance for bioterrorism. The increased interest in surveillance methodology in the US which has followed the 9/11 terrorist attack is notable. Kaufmann et al. (1997) stated that the delay of one day in detection and response to an epidemic due to a bioterrorist attack can result in the loss of thousands of lives and millions of dollars. In the US, several new types of data are now being collected, such as nurse hotline calls, poison center calls and over-the-counter sales of health remedies. New types of systems to analyze the data and meet the demand for surveillance have been developed recently. A review of some of these systems is given by Lober et al. (2002). The surveillance for bioterrorism can be done for example by public health surveillance or environment control.

Monitoring for detection of changes in public health is described for example by Sonesson and Bock (2003). In public health a quick detection is beneficial both at an individual level and to society. The timely detection of various types of adverse health events is an important issue. In the US there have been national conferences focused on the surveillance in public health. These conferences have been organized each of the last three years, and the CDC (Centers for Disease Control and Prevention) now has information on methodological issues of surveillance and displays an annotated bibliography on its website. The detection of outbreaks of infectious diseases such as SARS and Avian influenza (“bird flu”) has been of recent interest. The CDC website gives guidelines for **“**Influenza

A(H5N1) and SARS: Interim Recommendations for Enhanced U.S. Surveillance, Testing, and Infection Controls". Systems for data collection which can be used for surveillance have recently been built also in Asian countries (Chuc and Diwan (2003)). Monitoring of mortality rates in primary care in order to detect serious causes such as the serial killing by Shipman or less serious but important health care defects is treated for example by Aylin et al. (2003). Methods for post marketing surveillance of adverse effects of drugs are described for example by Andrew et al. (1996). The new systems for automatic registration of health events make new kinds of surveillance systems possible.

The technical advancements with continual registrations of the status of a patient for example in intensive care motivate new kinds of surveillance systems. One example is the surveillance of the fetal heart rate during labor described by Frisén (1992). An abnormality, caused for example by a lack of oxygen due to the wrapping of the umbilical cord around the neck of the fetus, might occur at any time. Detection has to take place as soon as possible after the event has occurred to ensure that a rescue action, such as a Caesarean section, is successful.

Needs for environmental control are described for example in the book edited by Barnett and Turkman (1993). Surveillance technique used for monitoring of biodiversity is described for example by Pettersson (1998). Järpe (2001) describes monitoring for detection of increased radiation caused by a possible nuclear plant disaster. The spatial clustering of various forms of cancer might be associated with environmental hazards. Warning systems for a natural disaster such as a tsunami are also of recent interest.

Applications in economics, especially the surveillance of business cycles, are treated for example by Andersson et al. (2004) and Andersson et al. (2005). Predicting the future state of the economy is important both to governments and to businesses. The changes between periods of recession and expansion can be predicted by monitoring leading indices. Systems for surveillance of threats to the financial stability of society have recently been of interest.

The aims in financial trading agree well with the demand on timeliness in surveillance. Shiryaev (2002) demonstrated that there might be an arbitrage opportunity when change points occur. Lam and Yam (1997) claimed to be the first to use methods for statistical surveillance as financial trading rules. Schmid and Tzotchev (2004) suggest an advanced system and state that the applications of surveillance methods in finance have been scarce up to now. Statistical surveillance as a financial trading strategy is described and reviewed for example by Bock et al. (2004)

3 THE SHEWHART METHOD AND THE STATISTICAL SURVEILLANCE PROBLEM

Shewhart (1931) presented the method which later got the name "the Shewhart method". This method is very simple and very popular in industrial quality control. It is still without doubt the most commonly used method of surveillance and now used for a variety of applications. Detailed descriptions are found in many textbooks like Wetherill and Brown (1991) and Ryan (2000).

We will first specify the general surveillance problem. We denote the process by $Y = \{Y(t): t = 1, 2, . . .\}$, where $Y(t)$ is the observation made at time t. The random process that determines the state of the system is denoted by $\mu(t)$. At each decision time, s, we want to discriminate between two states of the monitored system: the in-control state D(s) and the out-of-control state C(s). To do this, we can use only the observations up to time s,

$$Y_s = \{Y(t);t \leq s\} .$$

An alarm statistic, $p(Y_s)$, is based on these observations. With a control limit, G(s), we have the time of an alarm

$$t_A = \min\{s; p(Y_s)>G(s)\} .$$

Depending on the application, the time, τ, of the change can be regarded either as an unknown parameter or as a random variable with the probabilities $\pi(t) = P(\tau=t)$. How τ is regarded, govern the appropriate evaluation of the stochastic properties of methods. When τ is regarded as a random variable, the intensity, $\nu(t)$, of a change is defined as $\nu(t) = P(\tau=t|\tau \geq t)$, which is usually assumed to be constant over time.

Depending on the application, different types of in-control and out-of-control states are of interest. One interesting case is when $D(s) = \{\tau>s\}$ and $C(s) = \{\tau \leq s\}$.

For the Shewhart method, an alarm is triggered as soon as an observation deviates too much from the target. The observed value can be transformed so that Y is the deviation from the target. Only the last observation, Y(s), is considered in the Shewhart method. For the one-sided case an alarm is triggered at

$$t_A = \min\{s; Y(s) > G\},$$

where G is a constant. For the two-sided case the alarm statistic is |Y(s)|. The method can be regarded as performing repeated significance tests and the constant, G, is usually chosen to correspond to a certain significance level in the separate tests. The observation Y(s) could be a derived variable. Often the mean of several observations at the same time point is used. The Shewhart method is then referred to as the $\bar{X}$-chart.

The change to be detected and the stochastic properties of the process differ depending on the application. Several interesting cases will be described in Section 8. Until then, however, we will use a simple standard situation whenever a specification is necessary. This is in order to first concentrate on the general questions. Like in most studies in literature we have a step change, where a parameter changes from one constant level to another constant level. A shift in the mean of a normally distributed variable from an acceptable value μ^0 (say zero) to an unacceptable value μ^1, but otherwise iid, is used for specification. We have $\mu(t) = \mu^0$ for t= 1, . . , τ-1 and $\mu(t) = \mu^1$ for t= τ, τ+1 and onwards. For clarity, when suitable, standardization to $\mu^0=0$ and $\sigma=1$ is used and the size of the shift after standardization is denoted by μ. The case $\mu>0$ is described here. The case $\mu<0$ is treated in the same way.

4 EVALUATIONS OF SYSTEMS FOR SURVEILLANCE

Quick detection and few false alarms are good properties. When monitoring is used in practice, knowledge about the properties of the method is important. Otherwise it is hard to know to what degree an alarm indicates a change. In applied work evaluation by several measures may be necessary.

Using a constant probability of exceeding the alarm limit for each decision time means that we have a system of repeated significance tests. This may work well also as a system of surveillance and is often used. The Shewhart method has this property.

When we study the stochastic properties of a method for surveillance the stochastic framework is important. Different measures are in focus depending on this. It makes a difference whether the time, τ, of the change is regarded as an unknown parameter or as a random variable with the probabilities $\pi(t) = P(\tau=t)$. A third view is to regard these probabilities as priors for parameters in a Bayesian framework. Here, τ is regarded as an ordinary stochastic variable, and a frequentistic framework is used.

Different error rates and their implications for a system of decisions were discussed by Frisén and de Maré (1991). Illustrations of some of the measures mentioned below are found in the computer program by Frisén and Gottlow (2003). Formulas for numerical approximations of some of the measures are available in the literature.

4.1 False Alarms

The most commonly used measure of the false alarm rate is the average run length, $ARL^0 = E(t_A \mid D(\infty))$, where $D(\infty)$ denotes the case where no change ever occurs. The limit G for the Shewhart method for a fixed ARL^0 is calculated by the relation: $P(Y(s)>G \mid \mu(s)=\mu^0)=1/ARL^0$.

Another measure is the false alarm probability, PFA = $P(t_A<\tau)$. This is the probability that the alarm occurs before the change and is relevant if τ is regarded as a random variable. In theoretical work, the standard procedure is to assume that τ is geometrically distributed, implying a constant intensity. The relation between the PFA and the ARL^0 is different for different methods, as demonstrated by Frisén and Wessman (1999). Thus, comparisons between methods can give different results depending on whether the value of the PFA or the ARL^0 is fixed. The Shewhart method is favoured by a fixed ARL^0.

Chu et al. (1996) and others advocate monitoring methods which have a fixed (asymptotic) probability of any false alarm during an infinitely long surveillance period.

$$\alpha = \lim_{n\to\infty} P(t_A \le n \mid D(\infty) < 1 .$$

For some applications, this might be important because a strict significance test is the objective. In that case, ordinary statements for hypothesis testing about, for example, size and power can be made. However, the price for this additional feature is high as the detection ability of methods with this property declines rapidly with the value of time τ of the change. The expected delay of the detection of a change will be very large, as pointed out by Pollak and Siegmund (1975) and Frisén (2003). A demonstration of the drawbacks is also made by Bock (2004) by a simulation study.

4.2 Delay of the Alarm

The expected delay from the time of change, τ=t, to the time of alarm, t_A, is

$$ED(t) = E[\max (0, t_A-t) \mid \tau=t]$$

The ED(t) will typically tend to zero as t increases. However, the conditional expected delay for a specific change point τ

$$CED(t) = E[t_A-\tau \mid t_A \exists \tau=t] = ED(t) / P(t_A \exists t)$$

has a positive asymptote for most methods. Some extreme methods which have very good ARL properties, but for which CED(t) tends to infinity when t increases, are discussed by Frisén and Sonesson (2005b). The CED is constant for the Shewhart method. For other methods it is generally not the same for early changes as for late ones. The overall expected delay, with respect to the distribution of τ, is of interest if the application requires that τ should be regarded as a random variable. We have

$$ED = E[ED(\tau)],$$

where the expectation is with respect to the distribution of τ. A measure which is related to CED(t) and ED(t) is the Mean Time to Detection

$$MTD(t) = E[t_A - t \mid \tau = t].$$

This is used in the literature on signal detection (see for example Gustavsson (2000)). The expected value, with respect to the distribution of τ, differs from ED and there is no simple relation to the ED optimality described in Section 5.

The most commonly used measure of the delay is ARL^1 (see for example Page (1954) and Ryan (2000)). It is the average run length until detection of a true change (that occurred at the same time as the surveillance started). The part of the definition in the parenthesis is seldom spelled out. Note that

$$ARL^1 = ED(1)+1.$$

For some situations and methods the properties are about the same regardless of when the change occurs, but this is not always true, as illustrated by Frisén and Wessman (1999). Gan (1993) advocates that the median run length be used instead of the average. However, the main issue is that only the case τ=1 is considered.

Sometimes there is a limited amount of time available for rescue actions. The Probability of Successful Detection suggested by Frisén (1992) measures the probability of detection with a delay time no longer than a specified value, say d

$$PSD(d,t) = P(t_A - \tau < d \mid t_A \geq \tau = t).$$

This measure is a function of both the time of the change and the length of the interval in which the detection is defined as successful. It has been used for example by Petzold et al. (2004) in connection with a monitoring system for pregnancies. Even if there is no absolute limit for the detection time it may be useful to describe the ability to detect the change within a certain time limit. This has been done by Marshall et al. (2004) in connection with monitoring of health care quality.

4.3 Predictive Value

The appropriate action at an alarm differs depending on whether we are sure that there has been a change or we just have a low suspicion. The predicted value, PV, is the probability that a change has occurred when the surveillance method signals (Frisén (1992))

$$PV(t) = P(\tau \leq t_A \mid t_A = t).$$

Some methods have a constant PV. Others have a low PV at early times but a higher one later. In such cases, the early alarms will not motivate the same serious action as later alarms.

5 OPTIMALITY CRITERIA

In order to relate the Shewhart method to optimal followers, we will first describe and discuss some optimality criteria in surveillance. We will use the measures of the previous section to formulate some criteria and illustrate them by the standard case.

5.1 Minimal Expected Delay, ED

The minimal expected delay of a motivated alarm is a natural criterion. Shiryaev (1963) suggested a very general utility function. It was demonstrated by Frisén (2003) that the ED expression plays an important role. The gain of an alarm should be a linear function of the value of the delay, t_A-τ, and the loss associated with a false alarm can be an arbitrary function of the same difference. The utility can be expressed as U= E{u(τ, t_A)}, where

$$u(\tau,t_A)=\begin{cases} h(t_A-\tau) & \text{if } t_A<\tau \\ a_1(t_A-\tau)+a_2 & \text{else} \end{cases}$$

Girshick and Rubin (1952), motivate the utility function from incomes and costs at industrial production. When a false alarm causes the same cost of alerts and investigations irrespectively of how early the false alarm is, then the function h(t_A-τ) is a constant (say b), and we have

$$U= b\ P(t_A<\tau) + a_1\ ED + a_2\ .$$

Thus, we will have maximal utility if we have a minimal expected delay for a fixed probability of a false alarm. This is the ED criterion.

A variant of this utility function, based on exponential penalty, is suggested by Poor (1998) since an exponential penalty for the delay might be more relevant than a linear one for e.g. financial decisions. Beibel (2000) derives how his leads to a different optimal weighting of the observations.

The ED criterion has the advantage of including changes occurring at different time points. On the other hand, this might be experienced as a disadvantage when the intensity is unknown. In practice, however, there should be some knowledge available to influence the choice of method. Shiryaev (1963) treated the case of a constant intensity. Frisén and Wessman (1999) demonstrated that a non-informative intensity often is good enough. The risk of optimization for the wrong shift size, which the ED-criterion shares with other criteria, might be of greater concern. However, Andersson (2004) gave examples of large effect of misspecification of a not geometric distribution of the change point,.

5.2 Minimax

While the ED criterion uses the average delay after a change, it is also natural to look at the worst case and consider the minimax of the expected delay. It is related to the ED criterion as several possible change times are considered. However, instead of an expected value, which requires a distribution of the time of change, the worst value of CED(t) is used.

Lorden (1971) uses a still more pessimistic criterion, the "worst possible case" based not only on the worst value of the change time, but also on the worst possible outcome $Y_{\tau-1}$ before the change occurs. This criterion is pessimistic since it is based on the worst possible circumstances. Lorden (1971) studied the asymptotic properties but Moustakides (1986) demonstrated that the CUSUM method, described in Section 6, provides a solution to the finite criterion. The merits of studies of this criterion have been thoroughly discussed for example by Yashchin (1993) and Lai (1995). Shiryaev (1996) and Moustakides (2004) has studied minimax optimality of the CUSUM method for continous time.

Minimax methods based not on the delay, but on a function of the delay, has been studied by e.g. Poor (1998). This leads to different optimal weighting of the observations.

5.3 ARL

In statistical process control optimality is often stated as minimal ARL^1 for a fixed ARL^0. ARL^1 and ARL^0 are expectations under the assumption that the change affects either all or none of the observations.

Advantages of the ARL criterion are that it is very well known, widely used, simple and that no assumption of a distribution for the time of change is used. The ARL has been questioned as a formal optimality criterion by Frisén (2003) since some clearly inferior methods fulfill the criterion. For some (but not all) methods there is a correspondence to the minimax criterion.

5.4 Steady state ARL

The CED will for most (but not all) methods converge to a constant value. This asymptote is sometimes named the "steady state ARL" (Srivastava and Wu (1993)). This criterion concerns only the properties long after start. This is the opposite to the usual ARL which measures the behavior immediately after start. An advantage of this criterion is that it does not require any assumption of the intensity of the change.

5.5 Relations to Timeliness

For some applications optimality is not timeliness as such, but some measure related to timeliness. As an example, this is the case for financial transactions. The return of buying an asset at t = 0 and selling it at time t is

$$r(t) = Y(t) - Y(0)$$

where Y is a function of the price.

The expected return $E[r(t_A)]$ is maximized when $E[Y(t_A)]$ is maximized. This is achieved when we sell at the time, τ, of the peak. The return is measured along the price scale whereas the measures above are measured along the time scale. If $E[Y(t)]$ is linear on each side of the peak, then the relation is simple (Bock et al. (2004)).

6 THE FOLLOWERS AND THEIR OPTIMALITY PROPERTIES

The work by Shewhart has inspired much further research. Modifications motivated by applications to more complicated situations than the one described in Section 3 will be described in Section 8. Numerous modifications of the Shewhart method have been suggested even for the simple standard situation of Section 3. This simple situation is considered in this section whenever a specific distribution is needed. Here we will describe only some important and principally different methods. The Shewhart method utilizes only the last observation in spite of the fact that previous observations are available in the chart. Improvements in performance can be made by utilizing also other available observations.

The Shewhart method has no parameters to adjust except the alarm limit. Some followers are very flexible. The parameters can be chosen to make the method optimal for the specific conditions (for example the size of the change or the intensity of changes). In one way or another, many recent methods for surveillance are based on likelihood ratios. Thus, we will start by describing the full likelihood ratio method as it is a benchmark for other methods which, in different ways, are composed by partial likelihoods.

6.1 The Likelihood Ratio Method, LR

The full likelihood ratio for two states C and D is

$$f_{Ys}(y_s \mid C) / f_{Ys}(y_s \mid D).$$

At decision time s for a specified value, t, of the time of the change, this statistic is a weighted sum of the partial likelihoods

$$L(s,t) = f_{Ys}(Y_s \mid \tau = t) / f_{Ys}(Y_s \mid D(s)).$$

The alarm set consists of those Y_s for which the full likelihood ratio exceeds a limit. When the event to be detected at decision time s is $C(s)=\{\tau\leq s\}$ with the alternative $D(s)=\{\tau>s\}$, the time of an alarm for the LR method (Frisén and de Maré (1991)) is

$$t_A = \min\{s; \frac{f_{Y_s}(y_s|C(s))}{f_{Y_s}(y_s|D(s))} > \frac{P(\tau>s)}{P(\tau \leq s)} \cdot \frac{K}{1-K}\} = \min\{s; \sum_{t=1}^{s} w(s,t) \cdot L(s,t) > G(s)\}$$

where K is a constant, w(s,t) is a weight and G(s) is an alarm limit. The LR method is optimal with respect to the criterion of minimal expected delay and also a wider class of utility functions (Frisén and de Maré (1991)).

The time of an alarm can equivalently be written as the first time the posterior probability of state C exceeds a fixed level

$$t_A = \min\{s; P(C(s)| Y_s = y_s) > K\}.$$

Shiryaev (1963) and Roberts (1966) suggested a method, now called the Shiryaev-Roberts method, for which an alarm is triggered at the first time s, so that

$$\sum_{t=1}^{s} L(s,t) > G,$$

where G is a constant. The Shiryaev-Roberts method can be derived as the LR method with a non-informative prior for the distribution of τ. It can also be derived as the LR method when the change-point intensity, ν, tends to zero. A valuable property of the LR method is an approximately constant predictive value for the standard case described in Section 3 (Frisén and Wessman (1999)). Thus early and late alarms have the same interpretation of.

6.2 Window Methods

The Shewhart method is the most extreme window method since the window has size one. A special case of the likelihood ratio method with $C=\{\tau=s-d\}$ is

$$L(s, s-d) > G,$$

where G is a constant and d is a fixed window width. For the standard case of normally distributed variables described in Section 3 this will be a moving average. Thus optimal properties of error probabilities of the LR method carries over to the moving average method.

Sometimes, as in Lai (1998), advanced methods such as the GLR method are combined with a window technique in order to ease the computational burden.

6.3 Exponentially Weighted Moving Average Methods

A variant of a moving average method which does utilize all information is the EWMA method which was suggested by Roberts (1959). The alarm statistic is an exponentially weighted moving average,

$$Z_s = (1-\lambda)Z_{s-1}+\lambda Y(s),\ s=1, 2, ...$$

where $0<\lambda\leq 1$ and Z_0 is the target value, which is normalized to zero. If 8 is near zero, all observations have approximately the same weight. If $\lambda=1$, the EWMA method reduces to the Shewhart method. The asymptotic variant, EWMAa, will give an alarm at

$$t_A = \min\{s: Z_s > G\Phi_Z\},$$

where G is a constant and Φ_Z is the asymptotic standard deviation of Z_s. For the EWMAe version of the method, the exact standard deviation is used. For EWMAa the asymptotic is used. EWMAe can be regarded as a repeated significance test with a constant size. However, in a surveillance system the error probabilities are more complicated because of the repeated decisions. A comparison between the EWMAa and the EWMAe methods, for the standard case of Section 3, can be found in Sonesson (2003), where it is concluded that the EWMAa version is preferable for most cases. EWMAe give more frequent alarms at the first time point for the same ARL^0. CED(1) (and thus ARL^1) is best for EWMAe. However, since many of the alarms at time 1 are false for EWMAe, the predicted value of an alarm at $t_A=1$ is low. For larger values of t, EWMAa has better CED(t) than EWMAe.

The choice of λ is important, and the search for the optimal value of λ has been of great interest in literature. Small values of λ result in good ability to detect early changes, while larger values are necessary for changes that occur later. The statistic of the EWMA method is linear with respect to the observed values. It is of interest to compare the statistic with that of a linear approximation of the LR statistic for the standard situation of Section 3. The value of λ which makes the EWMA method identical to a linear approximation of the LR was determined by Frisén (2003). The conclusion was that when

$$\lambda = 1-\exp(-\mu^2/2)/(1-\nu),$$

the EWMA method is approximately optimal with respect to the minimal expected delay. This was confirmed by a simulation study by Frisén and Sonesson (2005b).

6.4 The CUSUM Method

The CUSUM method was first suggested by Page (1954). The alarm condition of the method can be expressed by the partial likelihood ratios as

$$t_A = \min\{s;\ \max(L(s, t);\ t=1, 2,.., s) > G\},$$

where G is a constant.

The CUSUM method satisfies the minimax criterion of optimality described in Section 5.

7 THE OPTIMALITY OF THE SHEWHART METHOD

The followers have many good properties. Still, the Shewhart method is by far the most commonly used one. Thus, it is of special interest to determine the situations for which this method is the best choice. The simple standard situation of Section 3 is considered in this section whenever a specific distribution is needed.

7.1 Immediate Detection

First we will see that the Shewhart method is a special case of the full likelihood ratio method which implies some optimality properties. Sometimes it is important to detect a change directly after it has happened. In such cases the state to be detected and the alternative at decision time s are

$$C(s) = \{ \tau = s\},\ D(s) = \{ \tau > s\}.$$

Frisén and de Maré (1991) proved that in the case of the simple situation of Section 3, with a shift in the mean of a normal distribution, the likelihood ratio method reduces to the Shewhart method. This can also be seen by expressing the likelihood ratio method as in Frisén (2003) with the alarm statistic

$$\sum_{t=1}^{s} w(s,t) \cdot L(s,t)$$

and the weights

$$w(s, t)=P(\tau=t)/P(\tau\leq s).$$

With C = { $\tau = s$} we have w(s, s)=1 and all other weights zero. The alarm set can be expressed by the condition

$$L(s, s) > G,$$

where G is a constant. This expression can be a used as a definition of a generalized Shewhart method. In the case of a shift in the mean of the normally distributed variable Y, this is the usual Shewhart expression with alarm at

$$t_A = \min\{s; Y(s) > G\}$$

Thus, for this situation the Shewhart method has the optimality properties of the full likelihood ratio method. The Shewhart method fulfils the ED criterion when C(s)={τ=s}. We also have minimal error probabilities for each decision time s (Frisén and de Maré (1991)).

7.2 Large Shifts

Frisén and Wessman (1999) proved, for the standard situation, that when the LR method is optimized for a shift size that tends to infinity, it converges to the Shewhart method. Thus, for large shifts, the Shewhart method has the same optimality as the LR method. It is thus ED-optimal for large shifts.

The Shewhart method can also be seen as a special case of the EWMA method. As mentioned in Section 6, the EWMA method is approximately ED-optimal if

$$\lambda = 1\text{-}\exp(-\mu^2/2)/(1-\nu).$$

This tends to 1 when μ tends to infinity. However, the EWMA method with $\lambda = 1$ is the Shewhart method. Also the relation to the EWMA method supports the ED-optimality of the Shewhart method for large shifts.

The relation to the CUSUM method gives further insight into the optimality of the Shewhart method. Frisén and Wessman (1999) proved, for the standard situation, that when the CUSUM method is optimized for a shift size that tends to infinity, it converges to the Shewhart method. Thus, for large shifts, the Shewhart method is minimax optimal.

7.3 Comments

The optimality results given above for large shifts and for immediate detection both support the same conclusion. The Shewhart method is to be chosen when a large change – which has to be detected immediately – is of main concern. In some applications, however, the two results may also be of interest separately.

For small shifts the situation is quite different, as demonstrated for example by the simulation study by Frisén and Sonesson (2005b). By the ARL- and ED-criteria, the Shewhart method performs poorly for small shifts.

The minimax criterion gives a relative favor to the Shewhart method. This is because the CED is constant. For other methods the CED is better than the worst CED for most of the possible change points, but this will not give any advantage with the minimax criterion.

8 THE SHEWHART METHOD AND ITS FOLLOWERS AT COMPLICATED SITUATIONS

When the states (between which the change occurs) are completely specified, the likelihood ratio method with its good optimality properties can be used. Pollak and Siegmund (1985) point out that the martingale property (for continuous time) of the Shiryaev-Roberts method makes it more suitable for adaptation to complicated problems than the CUSUM method. On the other hand, Lai (1995) and Lai (1998) point out that the good minimax properties of generalizations of the CUSUM method make the CUSUM suitable for complicated problems.

8.1 Gradual Changes

Most of the literature on surveillance treats the case of an abrupt change. However, in many cases the change is gradual, for example at the outbreak of a contagious disease or in the price of a financial asset. It may be hard to model exactly the shape of the rise and the decline, or even to estimate the baseline accurately. Then, the timely detection of a change in monotonicity is of interest. Frisén (2000) suggested surveillance that is not based on any parametric model but only on monotonicity restrictions. This surveillance method was described and evaluated by Andersson (2002) and Andersson (2004). It is developed for cyclical processes with the aim to detect a turn (peak or trough) as soon as possible. Instead of the full likelihood, the maximum likelihood ratios over the class of all monotonic or unimodal functions, respectively, are used. The maximum likelihood estimator of μ under the monotonicity restriction is described for example by Robertson et al. (1988). The maximum likelihood estimator under the unimodality restriction was given by Frisén (1986).

Arteaga and Ledolter (1997) compare several procedures with respect to ARL properties for several different monotonic changes. One of the methods suggested in that paper is a window method based on the likelihood ratio and isotonic regression techniques. In most cases, comparisons between window statistics are inefficient for the detection of gradual changes (Järpe (2000)). If a simple method is necessary, then the Shewhart method might be preferred at gradual changes. Yashchin (1993) discusses generalizations of the CUSUM and EWMA methods to detect both sudden and gradual changes.

8.2 Change between Unknown Parameters

The Shewhart method does not involve the size of the change as a parameter. This may be convenient since the value after a change is seldom known, but it can also be seen as a disadvantage since the possibility to optimize for a certain size is lost. However, false alarm properties will remain even if the size of the change is not known. The shift size is required only to design the method to be as powerful as possible.

To control false alarms is usually more important than to optimize the detection ability. Knowledge of the in-control parameters is important. Often these are estimated and used as a plug-in value in the method. The estimated baseline will affect the performance of the method. Kramer and Schmid (2000) study modifications and residuals for the Shewhart method to keep the false alarm rate under control in spite of the estimation of parameters. Albers and Kallenberg (2004) conclude that for the Shewhart method the sample size of the estimate has to be very large. Andersson et al. (2005) make the same conclusion for other methods.

One way to avoid the problem of unknown parameters is to transform the data to invariant statistics. Frisén (1992) and Sullivan and Jones (2002) use the deviation of each observation from the average of all previous ones. The comparison between two windows statistics is another usual transformation. Gordon and Pollak (1997) use invariant statistics combined with the Shiryaev-Roberts method to handle the case of an unknown pre-change mean of a normal distribution. Krieger et al. (2003) use invariant statistics combined with the CUSUM and Shiryaev-Roberts methods for surveillance of change in regression from an unknown pre-change mean.

When both the baseline and the increase at a change are unknown, we aim for the detection of a change to a stochastically larger distribution. Bell et al. (1994) suggested a nonparametric method geared to the exponential distribution. Baron (2000) proposes a method based on histogram density estimators to detect any change in distribution. The nonparametric maximum likelihood method above, designed for detection of a change from monotonicity, also avoids the problem of unknown parameters by using only the monotonicity properties.

The problem of unknown parameter values can be handled by a statistic, which is based on the maximum difference (measured for example by the likelihood ratio) between the baseline and the changed level. The GLR method (Lai (1995) and Lai (1998)) uses the maximum likelihood estimator. For the GLR method, the alarm statistic is formed by maximizing over possible values of the parameter value. Kulldorff (2001) used the same technique for detection of clustering in spatial patterns. Hawkins and Zamba (2005) use the GLR technique for monitoring variance. Lai (1998) proved a minimax result for a variant of GRL suitable for autocorrelated data. Lee and Lee (2004) study the limiting distribution when the maximum likelihood estimation is used in the CUSUM method.

The MLR method suggested for example by Pollak and Siegmund (1975) involves priors for the unknown parameters. Priors are also used by Radaelli (1996) for the Sets method and by Lai (1998) for the CUSUM method.

8.3 Events

Most of the theory on surveillance is derived for normally distributed data, but also discrete processes are of interest for many applications. Cardinal et al. (1999) describe the problems of using methods for continuous variables when studying observations which are based on count data. A bibliography of surveillance by the Shewhart method and other methods for attribute data is given by Woodall (1997).

In the study of events the case of a Poisson process is of special interest. Sometimes the exact time of the event is not known or convenience may require that only the numbers of events in fixed intervals are recorded. This number is usually assumed to be Poisson distributed. A direct analogue to the methods for the normal case is achieved by comparing the recorded number of events in each time period with the expected number. For the Shewhart method an alarm is given as soon as this difference is large. Lucas (1985) described the Poisson CUSUM method and Borror et al. (1998) the Poisson EWMA. Sonesson and Bock (2003) derived the likelihood ratio method.

When using the continuous exponentially distributed time between events, the Exponential CUSUM can be constructed (see for example Lucas (1985) and Lee and Lee (2004)). In contrast to the Poisson CUSUM, the Exponential CUSUM is not based on an initial reduction of the information. Exponential EWMA was described by Gan (1998). Sonesson and Bock (2003) derived the LR method based on the exponentially distributed time intervals between events. For the same situation, Kenett and Pollak (1996) gave the Shiryaev-Roberts method and made a comparison with the Shewhart method.

8.4 Multivariate Surveillance

Multivariate surveillance is needed in many areas. Sometimes measurements are obtained not only in time but also at various locations. The inferential problems involved in spatial surveillance are multivariate. Another example is the syndromic surveillance in public health where several symptoms are routinely recorded (see Lawson and Rodeiro (2004)). In several papers by Schmid and co-workers, multivariate economic surveillance is studied. An example is Sliwa and Schmid (2005) where the autocovariance and cross-covariance structures of financial assets in the Eastern European stock markets are monitored. Reviews of multivariate surveillance are given for example by Woodall and Montgomery (1999) and Sonesson and Frisén (2005).

There are some distinctly different approaches to handling multivariate surveillance. One approach is to reduce dimensionality. This can be done for example by using the principal components instead of the original variables, as proposed and discussed for example in Kourti and MacGregor (1996) and Scranton et al. (1996).

Another approach is to use scalar accumulation, where the components of the vector of observations are transformed into a scalar statistic for each time point before the accumulation over time. For example, the first step can be to calculate a spatial statistic for each time. Then, this statistic can be monitored in time by univariate surveillance. Wessman (1998) proved that when all the variables change at the same time, a sufficient reduction to univariate surveillance exists. Originally, the Hotelling T^2 statistic was used in a Shewhart approach. This is sometimes referred to as the Hotelling T^2 control chart. For multivariate binomial data, Lu et al. (1998) proposed a method based on a weighted sum of the number of adverse units for each of the components for each time point. The surveillance method used for this weighted sum was a Shewhart one. Other applications of Shewhart methods are made for example by Runger (1996) and Kang and Albin (2000).

A third approach is to use parallel univariate surveillance methods for each component variable and trigger an alarm if any of the univariate methods exceeds its limit in accordance with the union-intersection principle. Sometimes the Bonferroni method is used to control a false alarm error, see Alt (1985). As an example, in spatial surveillance one can make parallel surveillance of each location and make a general alarm when there is an alarm for any of the locations. This approach was used by Kulldorff (2001) for different cluster sizes and cluster locations.

Another approach is the vector accumulation approach. Here the alarm statistics of the parallel surveillance methods for each component variable are combined into a general alarm statistic. Lowry et al. (1992) proposed a multivariate extension of the univariate EWMA method, referred to as MEWMA, and demonstrated that this is an improvement on the Shewhart method applied to the χ^2-statistic. In both the vector and the scalar accumulation approaches, the correlations between the variables are used in the transformation. Rosolowski and Schmid (2003) and Sliwa and Schmid (2005) use the Mahalanobis distance to measure the distance between the target values and the actual values together with vector accumulation.

Still another approach is to handle the multivariate nature of the observational vector and the different time points simultaneously while aiming at satisfying some global optimality criterion. Järpe (1999) suggested an ED-optimal likelihood ratio surveillance method of clustering in a spatial log-linear model. If the alternatives are completely specified, general techniques to achieve minimax optimality can be found in Lai (1995).

Optimality is always complicated in multi-dimensional cases. No approach will be uniformly optimal for all kinds of changes. The multivariate methods can be evaluated by modifications of the measures and criteria described above. For example, Wessman (1999) suggested a generalization of the ARL measure to allow for the possibility of different change times for different variables. Control of the false discovery rate is of interest when conclusions are made about several variables and is used for example by Wong et al. (2003).

8.5 Dependency Structures

Four ways to treat the case of time dependent observations are discussed in Frisén (2003) and Frisén and Sonesson (2005a): 1) using the ordinary method and studying the robustness, 2) modifying the method with wider alarm limits based on the correct variance, 3) using the residuals from a time series model, or 4) deriving the method by optimality criteria for the full model.

Schmid and co-workers used all approaches with a focus on the modified and residual approaches. They studied the Shewhart and other methods for AR, ARCH and GARCH models (for example Schmid (1995), Schmid and Schöne (1997), Kramer and Schmid (1997), Severin and Schmid (1999), and Schipper and Schmid (2001)). In many of these papers, case studies are made on stock market data. A review is given in Knoth and Schmid (2004).

Liu and Tang (1996) suggested a nonparametric bootstrap-based generalization of the Shewhart method which does not require independent observations.

Petzold et al. (2004) derived the full likelihood ratio method for a MA model and compared it to a Shewhart method.

9 CONCLUDING REMARKS

The pioneering work by Walter Shewhart has inspired a new area of statistical development. The sequential character of the decision situation – without any fixed hypotheses – has been recognized in many applications. The availability of timely data and the need for timely decision-making have increased, and new applications have become important.

The issue of keeping calculations simple is less important now than it was when Shewhart suggested his method. Instead, the requirement of efficiency has increased. Optimality in surveillance is not a simple issue. There is a great difference between surveillance and hypothesis testing in this respect. Several criteria, which take care of the special features of a surveillance system, have been suggested. Some are suitable when the time of the change is considered as a random variable while others are suitable when it is considered as an unknown parameter. Some criteria require knowledge of a certain parameter such as the intensity of changes. The possibility to evaluate a method for a specific situation should not necessarily be seen as a disadvantage even though it is hard to exactly specify which situation is of most interest.

In spite of the good properties of the followers, the Shewhart method is probably still the one most commonly used in practice. The followers to the Shewhart method offer the possibility to optimize for the size of the shift and some also for the intensity of shifts. By studying how the Shewhart method is related to these optimal methods it can be concluded that the Shewhart method is well suited for quick detection of large shifts. As regards smaller shifts, different optimality criteria favor the Shewhart method to varying degrees. However, the Shewhart method is not the prime choice for small long-term changes.

Complicated situations, which are present in many important applications, have been studied for the Shewhart method and others.

10 ACKNOWLEDGEMENTS

The constructive suggestions and comments made by the referees are appreciated. This work has been partly supported by the Swedish Council for Research in the Humanities and Social Sciences, Grant F0473/2000.

11 REFERENCES

Albers, W. and Kallenberg, W. C. M. (2004) Estimation in Shewhart Control Charts: Effects and Corrections., *Metrika,* 59: 207-234.

Alt, F. B. (1985) Multivariate Quality Control, in *Encyclopedia of Statistical Science*, Vol. 6 Johnson, N. L. and Kotz, S. eds. pp. 110-122, New York: Wiley.

Andersson, E. (2002) Monitoring Cyclical Processes - a Nonparametric Approach, *Journal of Applied Statistics,* 29: 973-990.

Andersson, E. (2004) The Impact of Intensity in Surveillance of Cyclical Processes., *Communications in Statistics. Simulation and Computation,* 33: 889-913.

Andersson, E., Bock, D. and Frisén, M. (2004) Detection of Turning Points in Business Cycles, *Journal of Business Cycle Measurement and Analysis,* 1: 93-108.

Andersson, E., Bock, D. and Frisén, M. (2005) Statistical Surveillance of Cyclical Processes with Application to Turns in Business Cycles, *Journal of Forecasting,* To appear.

Andrew, J. E., Prescott, P., Smith, T. M., Inman, W. H. and Kubota, K. (1996) Testing for Adverse Reactions Using Prescription Event Monitoring., *Statistics in Medicine,* 15: 987-1002.

Arteaga, C. and Ledolter, J. (1997) Control Charts Based on Order-Restricted Tests, *Statistics & Probability Letters,* 32: 1-10.

Aylin, P., Best, N., Bottle, A. and Marshall, C. (2003) Following Shipman: A Pilot System for Monitoring Mortality Rates in Primary Care, *The Lancet,* 362: 485-491.

Banks, D. (1993) Is Industrial Statistics out of Control?, *Statistical Science,* 8: 356-409.

Barnett, V. and Turkman, K. F. (1993) *Statistics for the Environment,* Wiley.

Baron, M. (2000) Nonparametric Adaptive Change-Point Estimation and on-Line Detection, *Sequential Analysis,* 19: 1-23.

Beibel, M. (2000) A Note on Sequential Detection with Exponential Penalty for the Delay, *The Annals of Statistics,* 28: 1696-1701.

Bell, C., Gordon, L. and Pollak, M. (1994) An Efficient Nonparametric Detection Scheme and Its Application to Surveillance of a Bernoulli Process with Unknown Baseline, in *Change-Point Problems*, Carlstein, E., Muller, H.-G. and Siegmund, D. eds. pp. 7-27, Hayward, California: IMS Lecture Notes - Monograph Series.

Bock, D. (2004) Aspects on the Control of False Alarms in Statistical Surveillance and the Impact on the Return of Financial Decision Systems, *Submitted.*

Bock, D., Andersson, E. and Frisén, M. (2004) The Relation between Statistical Surveillance and Certain Decision Rules in Finance., *Submitted.*

Borror, C. M., Champ, C. W. and Rigdon, S. E. (1998) Poisson Ewma Control Charts, *Journal of Quality Technology,* 30: 352-361.

Cardinal, M., Roy, R. and Lambert, J. (1999) On the Application of Integer-Valued Time Series Models for the Analysis of Disease Incidence, *Statistics in Medicine,* 18: 2025-2039.

Chu, C.-S. J., Stinchcombe, M. and White, H. (1996) Monitoring Structural Change, *Econometrica,* 64: 1045-1065.

Chuc, N. T. K. and Diwan, V. K. (2003) Filabavi, a Demographic Surveillance Site, an Epidemiological Field Laboratory in Vietnam, *Scand J Public Health,* 31(Suppl. 62): 3–7.

Frisén, M. (1986) Unimodal Regression, *The Statistician,* 35: 479-485.

Frisén, M. (1992) Evaluations of Methods for Statistical Surveillance, *Statistics in Medicine,* 11: 1489-1502.

Frisén, M. (2000) Statistical Surveillance of Business Cycles. Research Report, Department of Statistics, Göteborg University.

Frisén, M. (2003) Statistical Surveillance. Optimality and Methods., *International Statistical Review,* 71: 403-434.

Frisén, M. and de Maré, J. (1991) Optimal Surveillance, *Biometrika,* 78: 271-80.
Frisén, M. and Gottlow, M. (2003) Graphical Evaluation of Statistical Surveillance. Research Report 2003:10, Statistical Research Unit, Göteborg University.
Frisén, M. and Sonesson, C. (2005a) Optimal Surveillance, in *Spatial Surveillance for Public Health*, Lawson, A. and Kleinman, K. eds., New York: Wiley.
Frisén, M. and Sonesson, C. (2005b) Optimal Surveillance Based on Exponentially Weighted Moving Averages, *Sequential Analysis,* To appear.
Frisén, M. and Wessman, P. (1999) Evaluations of Likelihood Ratio Methods for Surveillance. Differences and Robustness., *Communications in Statistics. Simulation and Computation,* 28: 597-622.
Gan, F. F. (1993) An Optimal Design of Ewma Control Charts Based on Median Run-Length, *Journal of Statistical Computation and Simulation,* 45: 169-184.
Gan, F. F. (1998) Designs of One- and Two-Sided Exponential Ewma Charts, *Journal of Quality Technology,* 30: 55-69.
Gordon, L. and Pollak, M. (1997) Average Run Length to False Alarm for Surveillance Schemes Designed with Partially Specified Pre-Change Distribution, *The Annals of Statistics,* 25: 1284-1310.
Gustavsson, F. (2000) *Adaptive Filtering and Change Detection,* Chichester: Wiley.
Hawkins, D. M. and Zamba, K. D. (2005) A Change-Point Model for a Shift in Variance, *Journal of Quality Technology,* 37: 21-31.
Järpe, E. (1999) Surveillance of the Interaction Parameter in the Ising Model, *Communications in Statistics. Theory and Methods,* 28: 3009-3025.
Järpe, E. (2000) On Univariate and Spatial Surveillance. Ph.D Thesis. Department of Statistics, Göteborg University.
Järpe, E. (2001) Surveillance, Environmental, in *Encyclopedia of Environmetrics*, El-Shaarawi, A. and Piegorsh, W. W. eds., Chichester: Wiley.
Kang, L. and Albin, S. L. (2000) On-Line Monitoring When the Process Yields a Linear Profile, *Journal of Quality Technology,* 32: 418-426.
Kaufmann, A. F., Meltzer, M. I. and Schmid, G. P. (1997) The Economic Impact of a Bioterrorist Attack: Are Prevention and Postattack Intervention Programs Justifiable?, *Emerging Infectious Diseases,* 3: 83-94.
Kenett, R. and Pollak, M. (1996) Data-Analytic Aspects of the Shiryaev-Roberts Control Chart: Surveillance of a Non-Homogeneous Poisson Process, *Journal of Applied Statistics,* 23: 125-137.
Knoth, S. and Schmid, W. (2004) Control Charts for Time Series: A Review., in *Frontiers of Statistical Quality Control.*, Vol. 7 Lenz, H.-J. and Wilrich, P.-T. eds. pp. 210-236, Heidelberg: Physika-Verlag.
Kourti, T. and MacGregor, J. F. (1996) Multivariate Spc Methods for Process and Product Monitoring, *Journal of Quality Technology,* 28: 409-428.
Kramer, H. and Schmid, W. (1997) Ewma Charts for Multivariate Time Series, *Sequential Analysis,* 16: 131-154.
Kramer, H. and Schmid, W. (2000) The Influence of Parameter Estimation on the Arl of Shewhart Type Charts for Time Series, *Statistical Papers, Berlin,* 41: 173-196.
Krieger, A. M., Pollak, M. and Yakir, B. (2003) Surveillance of a Simple Linear Regression, *Journal of the American Statistical Association,* 98: 456-469.

Kulldorff, M. (2001) Prospective Time Periodic Geographical Disease Surveillance Using a Scan Statistic, *Journal of the Royal Statistical Society A,* 164: 61-72.
Lai, T. L. (1995) Sequential Changepoint Detection in Quality-Control and Dynamical Systems, *Journal of the Royal Statistical Society B,* 57: 613-658.
Lai, T. L. (1998) Information Bounds and Quick Detection of Parameters in Stochastic Systems, *IEEE Transactions on Information Theory,* 44: 2917-2929.
Lai, T. L. (2001) Sequential Analysis: Some Classical Problems and New Challenges, *Statistica Sinica,* 11: 303-351.
Lai, T. L. (2004) Likelihood Ratio Identities and Their Applications to Sequential Analysis, *Sequential Analysis,* 23: 467-497.
Lam, K. and Yam, H. C. (1997) Cusum Techniques for Technical Trading in Financial Markets, *Financial Engineering and the Japanese Markets,* 4: 257-74.
Lawson, A. and Rodeiro, C. (2004) Developements in General and Syndromic Surveillance for Small Area Health Data, *Journal of Applied Statistics,* 31: 397-406.
Lee, S. and Lee, T. (2004) Cusum Test for Parameter Change Based on the Maximum Likelihood Estimator., *Sequential Analysis,* 23: 239-256.
Liu, R. Y. and Tang, J. (1996) Control Charts for Dependent and Independent Measurements Based on Bootstrap Methods, *Journal of the American Statistical Association,* 91: 1694-1707.
Lober, W. B., Karras, B. T., Wagner, M. M., Overhage, J. M., Davidson, A. J., Fraser, H., Trigg, L. J., Mandl, K. D., Espino, J. U. and Tsui, F. C. (2002) Roundtable on Bioterrorism Detection: Information System-Based Surveillance., *Journal of the American Medical Informatics Association,* 9: 105-115.
Lorden, G. (1971) Procedures for Reacting to a Change in Distribution, *Annals of Mathematical Statistics,* 42: 1897-1908.
Lowry, C. A., Woodall, W. H., Champ, C. W. and Rigdon, S. E. (1992) A Multivariate Exponentially Weighted Moving Average Control Chart., *Technometrics,* 34: 46-53.
Lu, X. S., Xie, M., Goh, T. N. and Lai, C. D. (1998) Control Chart for Multivariate Attribute Processes, *International Journal of Production Research,* 36: 3477-3489.
Lucas, J. M. (1985) Counted Data Cusum's, *Technometrics,* 27: 129-144.
Marshall, C., Best, N., Bottle, A. and Aylin, P. (2004) Statistical Issues in the Prospective Monitoring of Health Outcomes across Multiple Units, *Journal of the Royal Statistical Society A,* 167: 541-559.
Moustakides, G. V. (1986) Optimal Stopping Times for Detecting Changes in Distributions, *The Annals of Statistics,* 14: 1379-1387.
Moustakides, G. V. (2004) Optimality of the Cusum Procedure in Continuous Time, *The Annals of Statistics,* 32.
Page, E. S. (1954) Continuous Inspection Schemes, *Biometrika,* 41: 100-114.
Pettersson, M. (1998) Monitoring a Freshwater Fish Population: Statistical Surveillance of Biodiversity., *Environmetrics,* 9: 139-150.
Pettersson, M. (2004) Spc with Applications to Churn Management, *Quality and Reliability Engineering Internato,* 20: 397-406.

Petzold, M., Sonesson, C., Bergman, E. and Kieler, H. (2004) Surveillance in Longitudinal Models. Detection of Intrauterine Growth Retardation, *Biometrics,* 60: 1025-1033.

Pollak, M. and Siegmund, D. (1975) Approximations to the Expected Sample Size of Certain Sequential Tests, *The Annals of Statistics,* 3: 1267-1282.

Pollak, M. and Siegmund, D. (1985) A Diffusion Process and Its Applications to Detecting a Change in the Drift of Brownian Motion., *Biometrika,* 72: 267-280.

Poor, V. H. (1998) Quickest Detection with Exponential Penalty for Delay, *The Annals of Statistics,* 26: 2179-2205.

Radaelli, G. (1996) Detection of an Unknown Increase in the Rate of a Rare Event, *Journal of Applied Statistics,* 23: 105-113.

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

Roberts, S. W. (1966) A Comparison of Some Control Chart Procedures, *Technometrics,* 8: 411-430.

Robertson, T., Wright, F. T. and Dykstra, R. L. (1988) *Order Restricted Inference,* Chichester: Wiley.

Rosolowski, M. and Schmid, W. (2003) Ewma Charts for Monitoring the Mean and the Autocovariances of Stationary Gaussian Processes, *Sequential Analysis,* 22: 257-285.

Runger, G. C. (1996) Projections and the U^2 Chart for Multivariate Statistical Process Control, *Journal of Quality Technology,* 28: 313-319.

Ryan, T. P. (2000) *Statistical Methods for Quality Improvement,* New York: Wiley.

Schipper, S. and Schmid, W. (2001) Sequential Methods for Detecting Changes in the Variance of Economic Time Series, *Sequential Analysis,* 20: 235-262.

Schmid, W. (1995) On the Run Length of a Shewhart Chart for Correlated Data, *Statistical Papers, Berlin,* 36: 111-130.

Schmid, W. and Schöne, A. (1997) Some Properties of the Ewma Control Chart in the Presence of Autocorrelation, *The Annals of Statistics,* 25: 1277-1283.

Schmid, W. and Tzotchev, D. (2004) Statistical Surveillance of the Parameters of a One-Factor Cox-Ingersoll-Ross Model, *Sequential Analysis,* 23: 379-412.

Scranton, R. D., Runger, G. C., Keats, J. B. and Montgomery, D. C. (1996) Efficient Shift Detection Using Multivariate Exponentially-Weighted Moving Average Control Charts and Principal Components, *Quality and Reliability Engineering International,* 12: 165-171.

Severin, T. and Schmid, W. (1999) Monitoring Changes in Garch Processes, *Allgemeines Statistisches Archiv,* 83: 281-307.

Shewhart, W. A. (1931) *Economic Control of Quality of Manufactured Product,* London: MacMillan and Co.

Shiryaev, A. N. (1963) On Optimum Methods in Quickest Detection Problems, *Theory of Probability and its Applications,* 8: 22-46.

Shiryaev, A. N. (1996) Minimax Optimality of the Method of Cumulative Sum (Cusum) in the Case of Continuous Time., *Russian Math. Surveys,* 51: 750-751.

Shiryaev, A. N. (2002) Quickest Detection Problems in the Technical Analysis of Financial Data, in *Mathematical Finance - Bachelier Congress 2000*, Geman, H., Madan, D., Pliska, S. and Vorst, T. eds., Springer.

Sliwa, P. and Schmid, W. (2005) Monitoring the Cross-Covariances of a Multivariate Time Series, *Metrika,* 61: 89-115.

Sonesson, C. (2003) Evaluations of Some Exponentially Weighted Moving Average Methods, *Journal of Applied Statistics,* 30: 1115-1133.

Sonesson, C. and Bock, D. (2003) A Review and Discussion of Prospective Statistical Surveillance in Public Health, *Journal of the Royal Statistical Society A,* 166: 5-21.

Sonesson, C. and Frisén, M. (2005) Multivariate Surveillance., in *Spatial Surveillance for Public Health*, Lawson, A. and Kleinman, K. eds. pp. 169-186, New York: Wiley.

Srivastava, M. S. and Wu, Y. (1993) Comparison of Ewma, Cusum and Shiryaev-Roberts Procedures for Detecting a Shift in the Mean., *The Annals of Statistics,* 21: 645-670.

Stoumbos, Z. G., Reynolds Jr, M. R., Ryan, T. P. and Woodall, W. H. (2000) The State of Statistical Process Control as We Proceed into the 21st Century, *Journal of the American Statistical Association,* 95: 992-998.

Sullivan, J. H. and Jones, L. A. (2002) A Self-Starting Control Chart for Multivariate Individual Observations, *Technometrics,* 44: 24-33.

Tartakovsky, A. G. and Veeravalli, V. (2002) Change-Point Detection in Multichannel and Distributed Systems with Applications., in *Applications of Sequential Methodologies*, Mukhopadhyay, N., Datta, S. and Chattopadhyay, S. eds., New York: Marcel Dekker, Inc.

Wessman, P. (1998) Some Principles for Surveillance Adopted for Multivariate Processes with a Common Change Point, *Communications in Statistics. Theory and Methods,* 27: 1143-1161.

Wessman, P. (1999) Studies on the Surveillance of Univariate and Multivariate Processes. Ph D thesis, Department of Statistics, Göteborg University.

Wetherill, G. B. and Brown, D. W. (1991) *Statistical Process Control,* Chapman and Hall.

Wong, W. K., Moore, A., Cooper, G. and Wagner, M. (2003) Wsare: What's Strange About Recent Events?, *Journal of Urban Health,* 80: I66-I75.

Woodall, W. H. (1997) Control Charts Based on Attribute Data: Bibliography and Review, *Journal of Quality Technology,* 29: 172-183.

Woodall, W. H. and Montgomery, D. C. (1999) Research Issues and Ideas in Statistical Process Control, *Journal of Quality Technology,* 31: 376-386.

Yashchin, E. (1993) Statistical Control Schemes - Methods, Applications and Generalizations, *International Statistical Review,* 61: 41-66.

Zacks, S. (1983) Survey of Classical and Bayesian Approaches to the Change-Point Problem: Fixed Sample and Sequential Procedures of Testing and Estimation. In *Recent advances in statistics*, pp. 245-269.