



Research Report
Department of Statistics
Göteborg University
Sweden

Statistical surveillance

Exponentially weighted moving
average methods and public
health monitoring

Christian Sonesson

Research Report 2001:7
ISSN 0349-8034

Mailing address:
Dept of Statistics
P.O. Box 660
SE 405 30 Göteborg
Sweden

Fax
Nat: 031-773 12 74
Int: +46 31 773 12 74

Phone
Nat: 031-773 10 00
Int: +46 31 773 10 00

Home Page:
<http://www.stat.gu.se/stat>

Statistical Surveillance
Exponentially Weighted Moving Average Methods and
Public Health Monitoring

By CHRISTIAN SONESSON

Department of Statistics, Göteborg University, Sweden

The need for statistical surveillance has been noticed in many different areas and examples of applications include the detection of an increased incidence of a disease, the detection of an increased radiation level and the detection of a turning point in a leading index for a business cycle. In all cases, preventive actions are possible if the alarm is made early. If the change is detected too late, this can have severe consequences both at a personal level for affected individuals and to society as a whole. In these important situations we must evaluate the evidence value of the information we have about the process in order to guide us in the choice of making an alarm or not. The aim is to detect the change as quickly as possible and at the same time control the rate of false alarms. To do this efficiently we construct alarm systems using the available observations of the process, which are taken sequentially in time. (Note the difference from a test of one hypothesis.) The theory of statistical surveillance deals with the construction of alarm systems and the evaluation of such systems. This licentiate thesis consists of two papers with this common subject.

The first paper (1) deals with the properties of a special type of surveillance methods called EWMA methods. One attractive feature of EWMA methods is the easily interpretable alarms statistic, which is an exponentially weighted moving average of all available observations of the process. Several ways of constructing alarm limits to this statistic have previously been suggested in the literature. In this paper new types of evaluations of the performance of suggested variants are made and the results cast new light on both the merits of the variants and the optimality criteria commonly used. Methodological issues of general interest in the area of statistical surveillance are also treated, such as the definition of comparability between methods.

The second paper (2) deals with statistical surveillance in the area of public health. A critical review with emphasis on the inferential issues is made. The merits of different approaches are discussed and a new method is derived. Especially noticeable from the review is the lack of methods of surveillance of a spatial pattern, an area that includes many important applications, not only in public health.

Papers included in the licentiate thesis:

- (1) Sonesson, C. (2001) Evaluations of some exponentially weighted moving average methods, Research Report 2001:6, Department of Statistics, Göteborg University.
- (2) Sonesson, C and D. Bock (2001) Statistical issues in public health monitoring- A review and discussion, Research Report 2001:2, Department of Statistics, Göteborg University.

Acknowledgement:

My deepest thanks goes to my supervisor Professor Marianne Frisé for her constructive guidance and never ending enthusiasm during the work with this licentiate thesis; to David Bock for the fun times during our cooperation on one of the papers; to Eva Andersson, Hans van Houwelingen and Andrew Lawson for their helpful comments on earlier versions on the manuscripts; to Eric Järpe, Thomas Holgersson, Robert Jonsson and Kjell Pettersson for valuable discussions; to Catrin Eklöf-Olsson and Eva Bang for administrative help; to my fellow PhD students for their support and to all my friends at the department for contributing to a friendly and positive atmosphere.

Evaluations of Some Exponentially Weighted Moving Average Methods

By CHRISTIAN SONESSON

Department of Statistics, Göteborg University, Sweden

SUMMARY

Several versions of the EWMA (Exponentially Weighted Moving Average) method for monitoring a process with the aim of detecting a shift in the mean are studied both for the one-sided and the two-sided case. The effects of using barriers for the one-sided alarm statistic are also studied. One important issue is the effect of different types of alarm limits. Different measures of evaluation are considered such as the expected delay, the ARL^1 , the probability of successful detection and the predictive value of an alarm to give a broad picture of the features of the methods. Results are presented both for a fixed ARL^0 and a fixed probability of a false alarm. The differences highlight the essential problem of how to define comparability between surveillance methods. The results are from a large-scale simulation study. Special attention is given to the effect on the confidence in the final results by the stochastic variation in the calibration of the methods. It appears that important differences from an inferential point of view exist between the one- and two-sided versions of the methods. It is demonstrated that the method, usually considered as a convenient approximation, is to be preferred over the exact version in many respects.

Key Words: CHANGE POINT, DETECTION, EXPECTED DELAY, EWMA, OPTIMALITY, PROBABILITY OF SUCCESSFUL DETECTION, QUALITY CONTROL, SURVEILLANCE

1 INTRODUCTION

In many areas the problem of detecting a change in a stochastic process through sequential observations is important. Examples include an increased variation in an industrial production process, an increase in the mean radiation level or an increased cancer incidence. The general goal is to detect the change in the process, occurring at an unknown time point, as quickly as possible after it has occurred and at the same time control the rate of false alarms in order to be able to take appropriate actions. Common to the situations mentioned is that a decision of whether the change has occurred in the process or not has to be made sequentially, based on the data collected so far. This means that traditional hypothesis testing cannot be used, neither can sequential tests, since we cannot stop sampling in favour of the null hypothesis. Instead we have to use statistical surveillance methods.

In this paper, the case of a positive shift in the mean of a normal distribution from one constant level, μ_0 , to another constant level, μ , at a random change point τ is considered. Let the stochastic process under surveillance be denoted by $X = \{X(t); t = 1, 2, \dots\}$ where $X(t), t \geq 1$, are assumed to be conditionally independent, given the change point, with equal variance. Without loss of generality we take $\mu_0 = 0$ and $\sigma_X = 1$. Then, $X(t) \sim N(\mu(t), 1)$, where $\mu(t) = \mu \cdot I(\tau \leq t)$ and $I(\tau \leq t)$ is the indicator function taking the value 1 if $\tau \leq t$ and 0 otherwise. At each time point, s , we want to discriminate between two states of the process, the in control state, $D(s) = \{\tau > s\}$ and the out of control state $C(s) = \{\tau \leq s\}$. To do this, we use an alarm system which consists of two parts; an alarm statistic, $p(X_s)$, where $X_s = \{X(t); t \leq s\}$, and an alarm limit, $g(s)$. The time of an alarm, t_A , is

$$t_A = \min\{s; p(X_s) > g(s)\}.$$

There are several ways to construct the alarm system. In this paper we will study different kinds of EWMA methods, all with an alarm statistic, which is an exponentially weighted moving average of the observations. The use of this alarm statistic was first introduced in statistical process control by Roberts (1959). It can be used to detect a positive shift in the mean, or a negative, or either of them. If we are only considering shifts in one direction, this is referred to as the one-sided case contrary to the two-sided case, where we are interested in shifts in either of the two directions. Several ways of constructing alarm limits to this alarm statistic has been proposed in the literature. Our focus in this paper will be on the effect of the different alarm limits on the performance of the methods.

The paper is organized in the following way. In Section 2, the different types of EWMA methods included in the study are presented. In Section 3 different ways to achieve comparability of methods are discussed. The requirements on the number of replicates in the simulation study for sufficient accuracy of conclusion is also discussed in this section. In Section 4 the comparisons of the methods for the one-sided case are presented. The one-sided case is compared with the one-sided case with a barrier and the two-sided case in Section 5. Some important differences are examined. In Section 6, some concluding remarks are given.

2 DIFFERENT TYPES OF EWMA METHODS

The EWMA method has been widely studied in the literature, mostly for various shifts in the mean in a stochastic process of normally distributed variables (Roberts (1959); Robinson and Ho (1978); Crowder (1987); Crowder (1989); Lucas and Saccucci (1990); Srivastava and Wu (1993); Chandrasekaran, English, and Disney (1995); Srivastava and Wu (1997); Schöne, Schmid, and Knoth (1999); Steiner (1999); Chan and Zhang (2000) and Friséen and Sonesson (2001)). However, the EWMA method has also been studied in many other situations. The case of a shift in the variance was studied by Chang and Gan (1994), Crowder and Hamilton (1992) and Morais and Pacheco (1998). The case of a simultaneous shift in both the mean and the variance has been studied by Domangue and Patch (1991), Gan (1995) and Morais and Pacheco (2000). Also shifts in other types of processes have been studied. Gan (1998) studied methods to detect a shift in a parameter of an exponentially distributed variable. The case of shift in the mean of an autocorrelated process was studied by Lu and Reynolds (1999), and the robustness to non-normality by Borrór, Montgomery, and Runger (1999). Multivariate EWMA methods have been studied by Tsui and Woodall (1993) and Runger and Prabhu (1996). The EWMA statistic has been shown to be useful also in other situations such as forecasting in time series (Box, Jenkins, and MacGregor (1974)).

The EWMA methods are based on an exponentially moving average, Z_s , of all accumulated observations. The alarm statistic can equivalently be represented by the recursive formula

$$Z_s = (1 - \lambda)Z_{s-1} + \lambda X(s),$$

where the weight parameter $\lambda \in (0,1]$. If $\lambda = 1$ only the last observation is used in the alarm statistic and the smaller λ is, the more equally weighted are the observations. It is well known that the standard deviation of Z_s is an increasing function of the time s and converges as s tends to infinity. The convergence will be slower for small values of λ .

$$\lim_{s \rightarrow \infty} \sigma_{Z_s} = \lim_{s \rightarrow \infty} \sqrt{\frac{\lambda}{2-\lambda} [1 - (1-\lambda)^{2s}]} = \sqrt{\frac{\lambda}{2-\lambda}}$$

What differs between the EWMA methods is the way of constructing alarm limits $g(s)$, the starting value Z_0 and whether or not a barrier for the alarm statistic is used. A barrier, b , constitutes a boundary of the alarm statistic. The one-sided version with a barrier, to detect positive shifts, uses an alarm statistic of the form $\max(b, Z_s)$. In this paper the focus is to compare the two most common ways of constructing the limits, both based on the standard deviation of the alarm statistic. The first one, here named EWMAe, uses the exact standard deviation of Z_s when constructing the alarm limits. The time of an alarm, for the one-sided case, is

$$t_A = \min\{s; Z_s > L \cdot \sigma_{Z_s}\}.$$

Since the standard deviation is increasing in time, so is the alarm limit of the EWMAe method.

The second way of constructing the alarm limits, and by far the most studied in the literature, here named EWMAa, uses the asymptotic standard deviation of Z_s and the time of an alarm, for the one-sided case, is

$$t_A = \min\{s; Z_s > L \cdot \sigma_Z\}.$$

The EWMAa method was introduced by Roberts (1959) and most of the previous studies have been concerned with this method. Robinson and Ho (1978) used an Edgeworth expansion to get a recursive technique to evaluate the average run length both in control ($ARL^0 = E[t_A | \tau = \infty]$) and out of control ($ARL^1 = E[t_A | \tau = 1]$). This was done for both the one-sided and two-sided EWMAa in discrete time. Crowder (1987) used integral equations to evaluate the properties of the run length distribution in discrete time for the two-sided EWMAa. In Crowder (1989) optimal values of the weight parameter λ to minimize the ARL^1 for a fixed ARL^0 , together with a design strategy of the method, was proposed, based on the results in Crowder (1987). Lucas and Saccucci (1990) used another approach to find the optimal value of λ for the same situation as by Crowder (1989). The alarm statistic was represented by a continuous state Markov chain. The properties were evaluated by discretizing the infinite-state transition probability matrix. For a comparison between these two approaches, see Champ and Rigdon (1991). Srivastava and Wu (1993) examined the EWMAa method in continuous time by representing it as a diffusion process. This was done for the one-sided case. Optimal values of λ to minimize the stationary average delay time, SADT, for

a fixed value of ARL^0 , were derived. After correcting for the overshoot these results were applied to the two-sided case for discrete time by Srivastava and Wu (1997), where optimal values of λ were derived for a fixed ARL^0 .

Although the types of limits used in EWMAe and EWMAa are the most common, other types have been proposed in the literature. Lucas and Saccucci (1990) suggested the use of the EWMAa method with a head start, that is $Z_0 \neq 0$. This approach assures a fast initial response to start-up problems in the process. Steiner (1999) proposed another type of fast initial response for the EWMAe method, where the usual alarm limit was multiplied by an exponential function, which lowered the alarm limit in the start. In Friséen and Sonesson (2001) the alarm limits were chosen to make the EWMA method a good approximation of a linearized version of the likelihood-ratio method for the one-sided case (see also Friséen (1999)). The likelihood ratio method gives an alarm as soon as the posterior probability of the process being out of control exceeds a fixed value. The latter two papers, Friséen (1999) and Friséen and Sonesson (2001) discussed the minimal expected delay, ED, for a fixed probability of a false alarm. The use of barriers for the EWMA alarm statistic has previously been studied in Gan (1998), where evaluations of ARL^1 were made for different values of the barrier for a fixed value of ARL^0 for the EWMAa method. This was done in the case of exponentially distributed observations.

Two-sided versions of the EWMA methods can be constructed in different ways. The most common way is to use symmetrical control limits in which case the alarm limit is constructed in the same way as for the one-sided case but instead using $|Z_s|$ as the alarm statistic. This is the approach analyzed in Section 6. Non-symmetrical alarm limits, or two parallel one-sided versions with barriers using a lower (and upper) bound of the alarm statistic, are also possible.

3 COMPARABILITY BETWEEN METHODS

3.1 *Choice of measure of false alarms*

When evaluating the effectiveness of different types of alarm systems, one has to face a trade off between false alarms and short delay times for motivated alarms. The way to handle this is usually in the same way as in a hypothesis-testing situation, where the type 1 error is fixed and evaluations of the power is made for various situations. The translation to the surveillance situation has traditionally been to characterize the type 1 error by the in control average run length to a false alarm, denoted by ARL^0 . Then different types of methods have been compared for a

fixed value of the ARL^0 . Another way of characterizing the type 1 error is by the probability of a false alarm, denoted by $P(t_A < \tau)$.

$$P(t_A < \tau) = \sum_{t=1}^{\infty} P(\tau = t)P(t_A < \tau | \tau = t)$$

The difference between these approaches is discussed by Frisén (1999) and Frisén and Sonesson (2001). In this paper evaluations of the different types of EWMA methods are carried out for both approaches. Comparisons are made for a fixed ARL^0 of 50 and 100 and also for a fixed value of $P(t_A < \tau)$ when τ is assumed to be geometrically distributed ($P(\tau = t) = v(1-v)^{t-1}$ for $t=1,2,\dots$) with $v=0.05$ or 0.01 . The fixed values chosen for $P(t_A < \tau)$ corresponds to the values for the likelihood ratio method when $ARL^0=100$ as given by Frisén and Wessman (1999). In Section 4.1 the effects of the choice of approach will be further discussed. Results will be presented for values of λ in the interval $[0.001, 0.40]$. Small values of λ result in run length distributions with heavy right tails (for some of the situations studied here) and are therefore too computationally time consuming. Therefore only values of λ larger than or equal to 0.001 were chosen.

3.2 Effect of uncertainty in false alarms measure

In order to fix either the ARL^0 or the probability of a false alarm, one has to choose the constant L in the alarm limit. This can be done in several ways. In several of the earlier studies of EWMA methods the ARL^0 has been approximated as a function of L . The formulas achieved have then been used to determine the appropriate L to get a certain desired ARL^0 (Crowder (1989); Lucas and Saccucci (1990) and Srivastava and Wu (1997)). In this paper the appropriate value of L is determined by simulations. Whatever method used to determine L , careful attention to the closeness to the desired ARL^0 or the desired probability of a false alarm is needed.

The main focus in many studies, including this one, is to evaluate different methods with respect to some measure, θ . For comparability this is done under the restriction that another measure, A , has a specified value. The measure θ is usually a function of the out of control run length distribution, which depends on the constant L . The measure θ could be for example the ARL^1 for different values of λ for the EWMAa method under the restriction that A (ARL^0 or $P(t_A < \tau)$) has a specified value. Other examples of possible measures θ would be the conditional expected delay, $CED(t) = E[t_A - \tau | t_A \geq \tau = t]$, or the expected delay,

$ED = \sum_{t=1}^{\infty} P(\tau = t) \cdot CED(t)$. The aim is to study the accuracy of θ , but to do that, we start by studying the accuracy of A .

Assume that the evaluation of θ should be done for some desired value A^* of A . We aim to determine L^* such that $A(L^*) = A^*$. The procedure here to estimate L^* is to choose L_1, L_2, \dots, L_n and use simulations, with the same number m of replicates for each value of L to estimate the values of $A(L_1), A(L_2), \dots, A(L_n)$. We approximate $A(L)$ with a linear function locally, $\hat{A}(L) = \hat{a}_1 + \hat{b}_1 L$, and choose L' accordingly to give $\hat{A}(L') = A^*$. A confidence interval for L^* can be constructed by considering the test of $H_0 : A(L) = A^*$. The model for our observations is $A(L) = a_1 + b_1 L + \varepsilon$, where $\varepsilon \sim N(0, \frac{\sigma^2_{IRL}}{m})$ and σ^2_{IRL} is the variance of the in control run length, assumed to be constant locally within the range of values of L considered for the regression. First we estimate σ^2_{IRL} with such a precision that we are able to neglect the standard error of this estimate and proceed as if σ^2_{IRL} was known. For the estimation of a_1 and b_1 we use OLS. Then for the estimator $\hat{A}(L)$ we know that

$$E[\hat{A}(L)] = A(L)$$

$$V(\hat{A}(L)) = \frac{\sigma^2_{IRL}}{m} \left(\frac{1}{n} + \frac{(L - \bar{L})^2}{\sum_{i=1}^n (L_i - \bar{L})^2} \right).$$

To test the hypothesis $H_0 : A(L) = A^*$ we use the test statistic $K(L) = \frac{\hat{A}(L) - A^*}{\sqrt{V(\hat{A}(L))}} \sim N(0,1)$,

which can also be used for the construction of confidence intervals for L^* . We denote the p :th percentile in the $N(0,1)$ distribution by z_p . Then, included in a two-sided confidence interval $[L_l, L_u]$ for L^* of confidence level $(1-\alpha)$ will be those L for which $K^2(L) \leq z^2_{1-\alpha/2}$ where we have

$$L_l = L' - \frac{z_{1-\alpha/2}}{\hat{b}_1} \frac{\sigma_{IRL}}{m} \sqrt{\frac{1}{n} + \frac{(L_l - \bar{L})^2}{\sum_{i=1}^n (L_i - \bar{L})^2}}$$

$$L_u = L' + \frac{z_{1-\alpha/2}}{\hat{b}_1} \frac{\sigma_{IRL}}{m} \sqrt{\frac{1}{n} + \frac{(L_u - \bar{L})^2}{\sum_{i=1}^n (L_i - \bar{L})^2}}.$$

From the confidence interval $[L_l, L_u]$ for L^* we can also construct a confidence interval for $L'-L^*$ as

$$\left[-\frac{z_{1-\alpha/2} \sigma_{IRL}}{\hat{b}_1} \frac{\sigma_{IRL}}{m} \sqrt{\frac{1}{n} + \frac{(L_l - \bar{L})^2}{\sum_{i=1}^n (L_i - \bar{L})^2}}, \frac{z_{1-\alpha/2} \sigma_{IRL}}{\hat{b}_1} \frac{\sigma_{IRL}}{m} \sqrt{\frac{1}{n} + \frac{(L_u - \bar{L})^2}{\sum_{i=1}^n (L_i - \bar{L})^2}} \right].$$

What we are interested in is $\theta(L^*)$ and our aim is to construct confidence intervals for the difference $\theta(L') - \theta(L^*)$. However, we do only have an estimate, L' , of L^* . After choosing a constant L' , we have an estimate $\hat{\theta}(L')$ of $\theta(L')$, and knowledge about its stochastic properties for this fixed value of L .

To construct a confidence interval for $\theta(L') - \theta(L^*)$ we approximate $\theta(L)$ by a linear function, $\theta(L) = a_2 + b_2 L$ locally around L' . Simulations of $\theta(L)$ using values $L_1, L_2, \dots, L_{n'}$ of L around L' can achieve confidence intervals for b_2 . For each value of L_i , $i = 1, \dots, n'$, let m'_i be the number of replicates that the estimate $\hat{\theta}(L_i)$ is based on. (For several possible θ , for example the $CED(t)$, the evaluation is based on the alarms for which $t_A \geq \tau$. The reason for the number of replicates to be unequal is due to the difficulty of dimensioning the simulations to achieve the same number of alarms at or after the change point for each value of L . Note though that for $\tau = 1$ this problem does not arise.) The regression parameters a_2 and b_2 will be estimated using WLS since the number of replicates for each value of L_i , $i = 1, \dots, n'$ is not the same. Our observations follow the model $\theta(L_i) = a_2 + b_2 L_i + \delta_i$, where $\delta_i \sim N(0, \frac{\sigma^2_{ORL}}{m'_i})$ and

σ^2_{ORL} is the variance of the out of control run length (only considering the alarms for which $t_A \geq \tau$). Also in this case we regard σ^2_{ORL} as known after estimating it with high precision. The WLS estimate of b_2 has the properties

$$E[\hat{b}_2] = b_2$$

$$V(\hat{b}_2) = \sigma^2_{ORL} \frac{n'}{n' \sum_{i=1}^{n'} m'_i L_i^2 - (\sum_{i=1}^{n'} \sqrt{m'_i} L_i)^2}.$$

We can then construct confidence intervals for b_2 using that $\frac{\hat{b}_2 - b_2}{\sqrt{V(\hat{b}_2)}} \sim N(0,1)$.

Let the constructed confidence interval, Λ , for $(L' - L^*)$ be of confidence $(1 - \alpha_1)$ and the constructed confidence interval, B , for b_2 be of confidence $(1 - \alpha_2)$. Then we can combine these intervals to construct a confidence interval of confidence at least $(1 - \alpha_1)(1 - \alpha_2)$ for $\theta(L') - \theta(L^*) = b_2(L' - L^*)$ taking $\min\{b \cdot l, b \in B, l \in \Lambda\}$ to be the lower limit and $\max\{b \cdot l, b \in B, l \in \Lambda\}$ to be the upper.

Confidence intervals for $\theta(L') - \theta(L^*)$ constructed in this way for some cases studied in the paper can be found in Table 1. Included are those situations where the variances in the in control run length distributions are the largest. These confidence intervals indicate that the determination of L' is good enough to guarantee reliable comparisons between the methods. The conclusion is that the numbers of replicates in the simulations are enough (but not unnecessary) for the present purposes.

Table 1. Confidence intervals for $CED(L') - CED(L^*)$ when $\mu = 1$, $ARL^0 = 50$, $n \geq 10$, $n' = 10$, $m = 500000$. Level of confidence = 0.9025.

Method	$\tau = 1$ ($m' = 1000000$)		$\tau = 20$ ($m' \geq 143500$)	
	Lower limit	Upper limit	Lower limit	Upper limit
EWMAa(0.001), 1 sided	-0.0008	0.0012	-0.0008	0.0013
EWMAe(0.001), 1 sided	-0.0008	0.0008	-0.0055	0.0056
EWMAa(0.001), 2 sided	-0.0009	0.0029	-0.0009	0.0029
EWMAe(0.001), 2 sided	-0.0023	0.0023	-0.0044	0.0045

4 COMPARISONS FOR THE ONE-SIDED CASE

We start by the comparisons between the EWMAa and the EWMAe methods for the one-sided case. In Section 5, the focus will be on the difference between constructing the methods for the one-sided case, the one-sided case with a barrier and two-sided case. Our main interest is to see the effect of the different alarm limits on the performance of the methods, both when the process is in- and out of control.

4.1 *On the difference between a fixed ARL^0 and a fixed probability of a false alarm*

The only thing that differs between the EWMAa and the EWMAe methods is the alarm limits. As already mentioned, the variance of the alarm statistic is an increasing function of time and converges as the time tends to infinity. This means that in order to fix the in control run length or the probability of a false alarm, the value of L in the alarm limit will be larger for the EWMAe than for EWMAa method, for a fixed value of λ . Since the convergence is slower for small values of λ the difference in L will be most pronounced for small values of λ . The difference in the alarm limits between the methods will effect their relative performances at different time points, both with respect to their false alarm distributions as well as the ability to detect the change at different time points as will be shown later.

An important thing is to notice the difference between using a fixed ARL^0 compared with a fixed probability of a false alarm. Since the in control run length distributions are different, an equal ARL^0 does not imply equal probability of a false alarm and vice versa. In Figure 1 the values of ARL^0 for a fixed value of the false alarm probability are presented.

For moderate values of λ , the value of ARL^0 is fairly constant if we have fixed the probability of a false alarm for both methods. There is only a slight difference between the EWMAa and EWMAe methods for these values of λ . However the required ARL^0 to fix the probability of a false alarm is much larger for smaller values of λ . This means a larger L for the case of a fixed probability of a false alarm for both methods. For small values of λ there is a large difference between the methods. We will therefore expect both the in control as well as the out of control performance to differ depending on the choice of a fixed ARL^0 or a fixed probability of a false alarm. We also expect different results for different values of λ . The difference in performance depends on our choice of ν . This can be seen comparing Figure 1a and 1b.

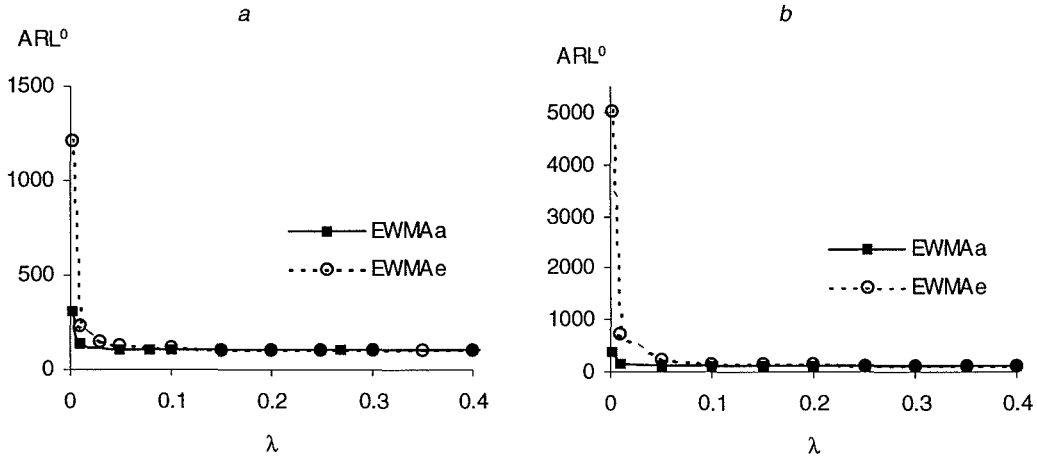


FIG 1: ARL^0 as a function of λ .
a. $P(t_A < \tau) = 0.4877$ for $\nu = 0.01$, b. $P(t_A < \tau) = 0.1326$ for $\nu = 0.05$.

The difference in ARL^0 for a fixed $P(t_A < \tau)$ reflects the importance of defining what is meant by comparability between methods of surveillance. The choice of fixing ARL^0 or $P(t_A < \tau)$ and at what level will affect the constants in the alarm limits. For the situations of a fixed $P(t_A < \tau)$ studied here, and shown in Figure 1, the adjustments needed in the value of L (measured as the increase in the ARL^0) when choosing a fixed $P(t_A < \tau)$ compared with a fixed value of $ARL^0 = 100$, is larger for EWMAe than for EWMAa and larger for small than for large values of λ . The values of $P(t_A < \tau)$, for which the methods have been fixed, are the ones, which result in a value of 100 for the ARL^0 for the likelihood ratio method (Frisén and Wessman (1999)), which can be used as a benchmark. From Figure 1, we can expect that the EWMAa method will perform relatively better than EWMAe and large values relatively better than small values of λ if we chose to fix $P(t_A < \tau)$ instead of ARL^0 . If this pattern is true in general with other values of $P(t_A < \tau)$, or another distribution of τ , is not examined here. The reason for expressing the difference between the cases in terms of the increase in ARL^0 in the case of a fixed $P(t_A < \tau)$ here is the fact that ARL^0 is the conventional measure of in control behavior and results expressed in this form can be easily interpreted. We can also consider how the alarm probabilities at different time points are affected. This comparison can be found in Figure 2 and 3.

In some cases there has been argued that ARL^0 is the only necessary in control characteristic with the motivation that the in control run length distribution is approximately geometric. The results here indicate that this is not the case, as will be seen in the figures displaying the in control run length distributions. What is clear though is that the difference between the methods in both in

control features as well as detection abilities will depend on our choice of in control characteristic. The most common way is to fix the ARL^0 . However, one should be aware of the consequences of such a choice.

4.2 In control features for the one-sided case

No matter whether the ARL^0 or the probability of a false alarm is fixed, the alarm limit for EWMAe will be lower than the one for EWMAa for early, but larger for later time points, for a fixed value of λ , due to the difference in time dependency of the alarm limits. The time point at which the alarm limits cross is later for small values of λ reflecting the slow convergence of the variance.

In Figure 2 we can see the effect of the different alarm limits on the in control run length distributions when $ARL^0=100$. The EWMAe method has a higher probability of a false alarm than the EWMAa method for a fixed value of λ at early time points since the alarm limits are lower. We can also notice a deviation from a geometric distribution, especially for the EWMAa method.

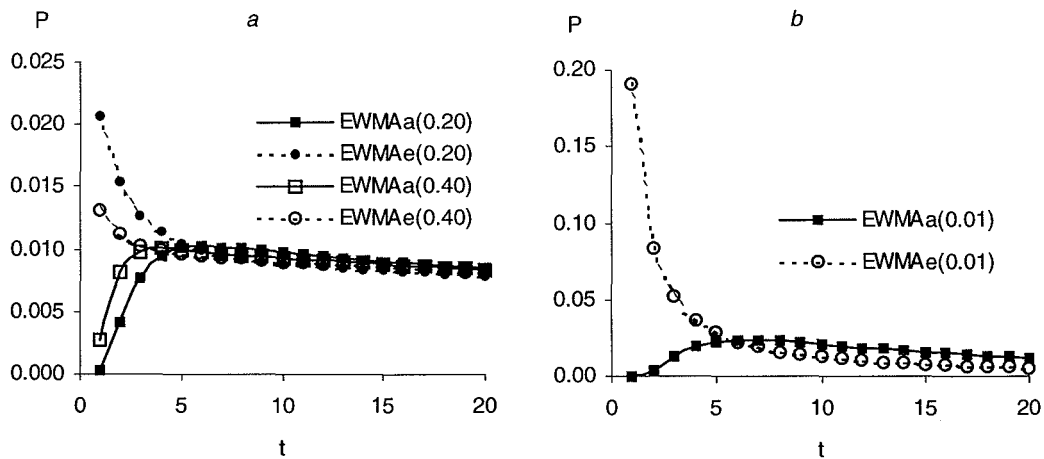


FIG 2: In control run length distribution when $ARL^0 = 100$. In the figure, the value of λ is indicated in the parentheses.

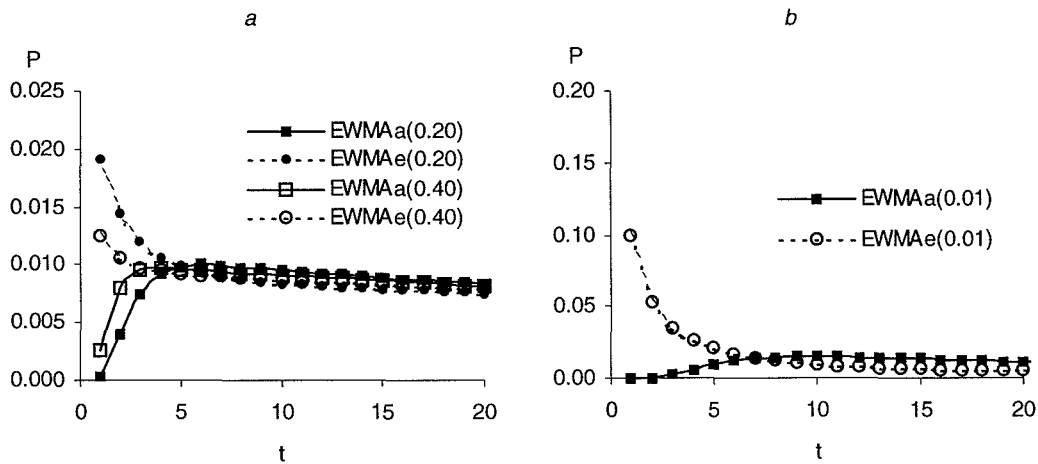


FIG 3: In control run length distribution when $P(t_A < \tau) = 0.4877$ for $\nu = 0.01$.

The in control run length distribution, when instead the probability of a false alarm is fixed, is illustrated in Figure 3 for the case when $\nu = 0.01$. We note especially the decreased probability of a false alarm at the first time point, comparing Figure 2b with Figure 3b, for the EWMAe when $\lambda = 0.01$ as a result of an increased value of L .

4.3 Detection of a true change for the one-sided case

There are several ways suggested in the literature to compare the performances of different surveillance methods with respect to true changes in the process. In the area of statistical process control the focus has traditionally been on the ARL^1 with optimality defined as minimal ARL^1 for a fixed value of ARL^0 . This criterion has also been the common one when designing the EWMA methods (Crowder (1989); Lucas and Saccucci (1990) and Srivastava and Wu (1997)). However, as an optimality criterion, this is not without critics (Frisén (1999) and Frisé and Sonesson (2001)). Gan (1993) considered instead the median out of control run length due to the skewed run length distributions. However, also in this case only shifts at the first time point were considered. This approach might be reasonable in an industrial manufacturing process, where one suspects various start-up problems. However, in the overwhelming majority of applications the possibility of later shifts should also be taken into account. One example of this is the monitoring of the foetal heart rate during labour (Frisén (1992)), where the foetus can suffer from a lack of oxygen. This can happen at any time point during the labour, which normally takes many hours, and thus in this case we must also take into account possible late shifts. Other examples include the surveillance of radiation levels (Järpe (2000)) and the surveillance of diseases (Sonesson and Bock (2001)). Since the false alarm probabilities at different time points differs between the methods, so will the

detection ability of a true change. Only considering shifts at the first time point will favour EWMAe over EWMAa with respect to the detection ability. This is because the false alarm distributions are different, as could be seen in Figure 2 and 3, where EWMAe gives more false alarms at early time points than EWMAa for the same value of λ .

In this paper we will use several kinds of measurements in order to get a broad picture of the ability of the methods to detect a true shift. Here we will consider the ARL^1 , the conditional expected delay, the expected delay, the probability of successful detection and the predictive value of an alarm.

4.3.1 A change at the first time-point for the one-sided case

As expected, due to the difference in the false alarm distributions, the EWMAe method has shorter ARL^1 compared with the EWMAa method for the same value of λ as can be seen in Figure 4, where $\mu = 1$ for $ARL^0=50$ and 100. Results for $\mu = 0.5$ can be found in Section 5.

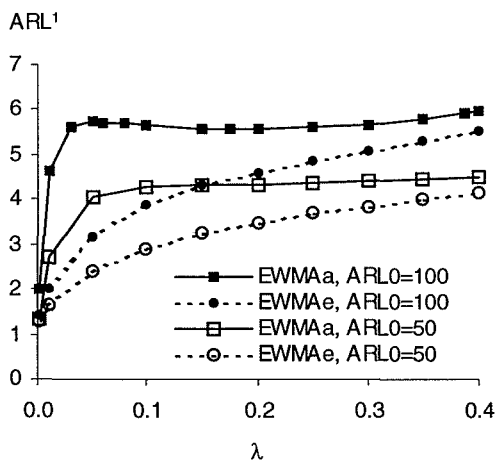


FIG 4: ARL^1 as a function of λ for $\mu = 1$ when $ARL^0 = 50$ and $ARL^0 = 100$.

Figure 4 indicates that in order to minimize the ARL^1 for a fixed value of the ARL^0 λ should approach zero (although the studied values of λ only covered the interval $[0.001, 0.40]$) both for EWMAa and EWMAe independently of the value of ARL^0 . Then, old and new observations have the same weight in the alarm statistic, and thus one of the necessary conditions given in Frisé (1999) for minimizing the ARL^1 is fulfilled. This will further be elaborated in Section 5, when comparing the results for the one-sided case with the two-sided case.

4.3.2 Detection of later changes for the one-sided case

The ability to detect changes at later time points, t , can be evaluated with respect to the conditional expected delay, $CED(t) = E[t_A - \tau | t_A \geq \tau = t]$. In the case of a change at the first time point, $CED(1) = ARL^1 - 1$. By considering the conditional expected delay, we are no longer limited to changes occurring at the first time point.

In Figure 5, the characteristics of the CED as a function of the time of the change can be examined. The EWMAe method has low delay times if the change occurs in the first time points. However the delay time increases with the time of the change. The results in Figure 5 clearly indicates that it is not enough only to consider changes at the first time point when evaluating EWMA methods since the time point of the change plays a crucial role in the detection ability. The increase in the observed values of CED is the largest for small values of λ . This holds both for the EWMAa and the EWMAe method and is the price to pay for the low delay times for early changes.

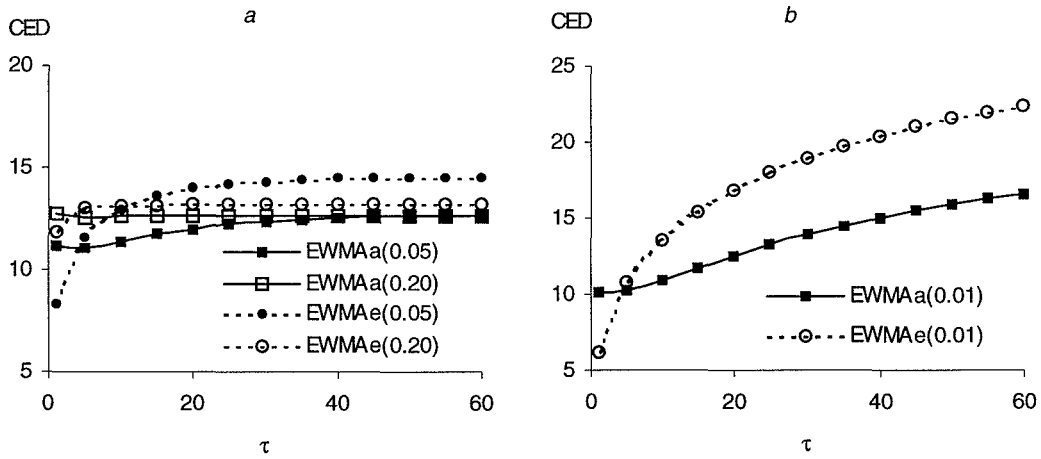


FIG 5: CED as a function of τ for $\mu = 0.5$ when $P(t_A < \tau) = 0.4877$ for $\nu = 0.01$.

To summarize the CED values for different time points different approaches can be taken. Two ways will be described here. The first one focuses on minimax properties of the methods with respect to the CED for various time points. The other way is to average the CED with respect to the distribution of the change-point τ .

Minimax criteria can be defined in various ways. Focus is on the maximum value of the conditional expected delay, which should be minimized. This type of evaluation has been studied

extensively in the literature (Lorden (1971), Pol lak (1985), Moustakides (1986)). One surveillance method known to possess optimality characteristics when the minimax criterion is expressed as minimal maximal conditional expected delay with respect to τ and the worst possible outcome of $X_{\tau-1}$ for a fixed value of the ARL^0 is the CUSUM method (Moustakides (1986)). In general, the CUSUM method is defined by the stopping rule

$$t_A = \min(s; w_s - \min_{0 \leq j \leq s} w_j > K),$$

where $w_s = \sum_{t=1}^s y_t$, $y_t = \log(f_1(X(t))/f_0(X(t)))$ and f_0 and f_1 denotes the in control and out of control distributions.

For the case studied in this paper, the time of an alarm for the CUSUM method can also be written recursively as

$$t_A = \min\{s; S_s > h\},$$

where $S_t = \max(0, S_{t-1} + X(t) - \mu/2)$ and $S_0 = 0$. Thus, the alarm statistic for the CUSUM method has a lower boundary of zero. For EWMA based methods, not much attention has been drawn to minimax evaluations. In Gan (1995) and Gan (1998) this has been discussed for the detection of different types of changes in exponentially distributed variables.

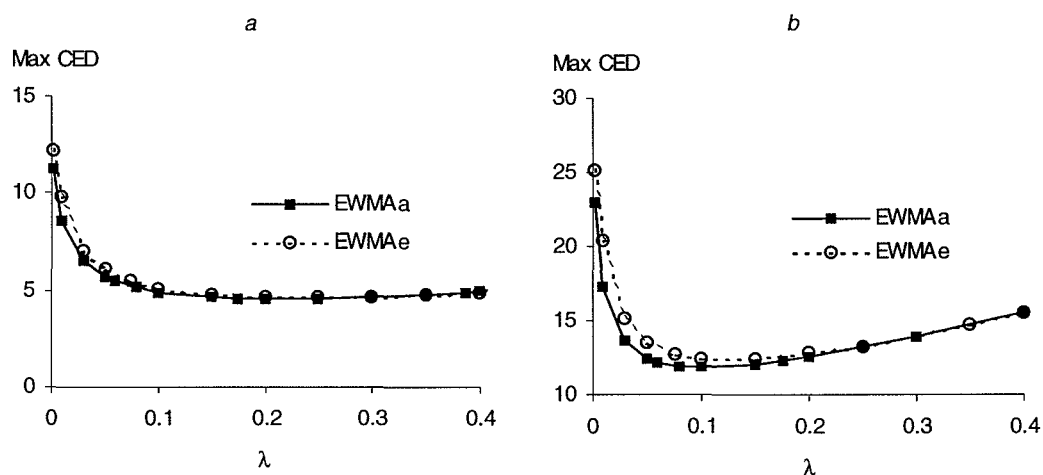


FIG 6: Maximal CED as a function of λ when $ARL^0=100$.

a. $\mu = 1$, b. $\mu = 0.5$.

In Figure 6 we see the effect of the different alarm limits on the maximum value of CED (with respect to the time point of the change) for a fixed value of the $ARL^0=100$. In the case when $\lambda \leq 0.01$ (where $CED(t)$ is still increasing with t when $t=100$) the values of $CED(100)$ are presented. In this sense, the EWMAa method is superior to the EWMAe. However, the difference between the methods is smaller if μ is large.

When summarizing the conditionally expected delay at different time points with respect to the distribution of the change-point τ , the standard procedure is to assume that τ is geometrically distributed with parameter ν . This implies a constant intensity of a shift, $\nu = P(\tau = t | \tau > t - 1)$. However, this assumption might be questioned in many applications. Assuming a distribution for τ , the expected delay is defined as

$$ED = E_{\tau} [t_A - \tau | t_A \geq \tau] = \sum_{t=1}^{\infty} P(\tau = t) \cdot CED(t).$$

The use of minimal expected delay for a fixed probability of a false alarm has been suggested as an optimality criteria and leads to the likelihood ratio method (Frisén and de Maré (1991)). This method is equivalent of making an alarm as soon as the posterior probability of an alarm exceeds a fixed value. The assumption made with respect to ν will determine the method and parameters which minimize the expected delay in the class of EWMA methods. If ν is large, methods with low $CED(t)$ for early time points will be preferable. If on the other hand ν is close to zero, methods with low values of $CED(t)$ at late time points will be preferred. Specifically, the value of ν will determine which of the EWMAa or EWMAe alarm limits that will be preferable as well as what value of λ that will minimize the expected delay. The expected delay can also be considered if the ARL^0 is fixed as is done in Figure 7.

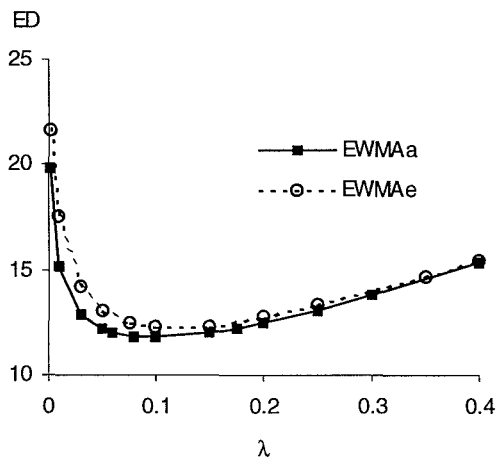


FIG 7: $ED(\nu=0.01)$ as a function of λ for $\mu=0.5$ when $ARL_0=100$.

For the case illustrated in Figure 7, the EWMAa method has lower expected delay than the EWMAe method. However, if ν is large enough, the EWMAe method has lower expected delay for all cases studied here. Consider for example the case when $\nu=1$, in which case only a shift at the first time point will be of interest and $ED = ARL^1 - 1$ and then EWMAe will have a lower expected delay as could be seen in Figure 4. However, a value of $\nu=0.01$ is not large enough for the EWMAe method to have a lower expected delay than the EWMAa method. In the majority of applications we expect that the EWMAa method will be preferable to the EWMAe method with respect to the expected delay. The suggested versions of the alarm limits to insure a fast initial response to start-up problems (Lucas and Saccucci (1990) and Steiner (1999)) can be expected to have even larger expected delay than the EWMAe method for cases with small values of ν . The difference between the EWMAa and the EWMAe method is larger, in absolute value, for a smaller change (not illustrated). This is in accordance with results in Frisén and Wessman (1999) (in that case regarding likelihood-based methods) that methods are alike if the change is large.

If instead we use a fixed value of the probability of a false alarm, the difference between the EWMAa and the EWMAe method is larger than for the case of a fixed ARL^0 , which is presented in Figure 8.

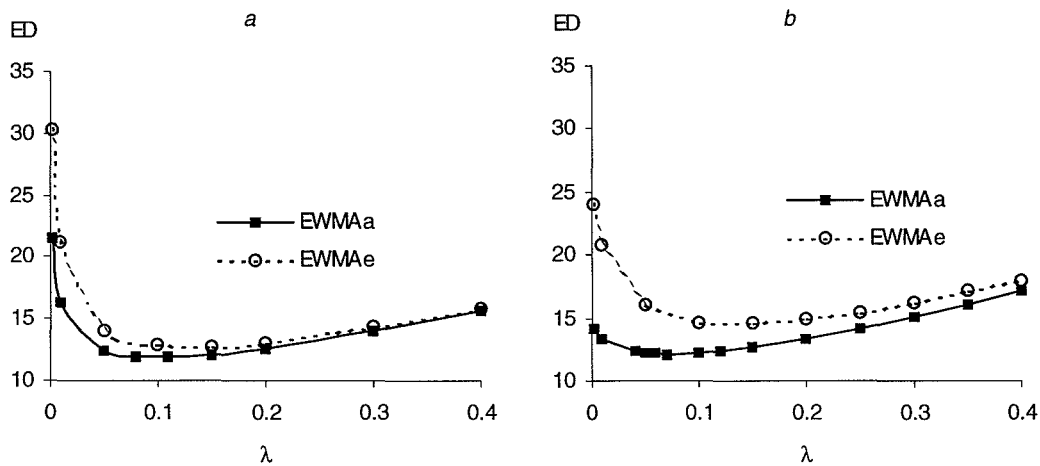


FIG 8:

a. $ED(v=0.01)$ as a function of λ for $\mu=0.5$ when $P(t_A < \tau) = 0.4877$ for $v=0.01$.

b. $ED(v=0.05)$ as a function of λ for $\mu=0.5$ when $P(t_A < \tau) = 0.1326$ for $v=0.05$.

Common to the measures of performance for detection of a true shift considered so far (ARL^1 , CED and ED) is that they focus on the mean of the out of control run length distribution conditioned on the time point of the shift. However, as in the in control case discussed previously, the mean does not account for all information in the out of control run length distribution. One alternative measure is the median delay time, considered in Gan (1993) for a shift at the first time point.

In some cases, the application considered calls for other types of evaluations than the mean delay time, for example the case where a limited time for actions exists. An example is the outbreak of an infectious disease where an epidemic starts if no actions are taken, or the case of surveillance of a foetus heart rate. For those cases, the expected value of the delay is of less interest. Instead the probability of successful detection, PSD, defined as the probability of detecting a change within a certain time interval, d , after a true change, is more important (Frisén (1992)).

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

The PSD is thus a function both of the time of the change and the length of the interval in which the detection is defined as successful. In Figure 9, the PSD is presented as a function of d for a fixed value of τ for the case of a fixed probability of a false alarm when $v=0.01$.

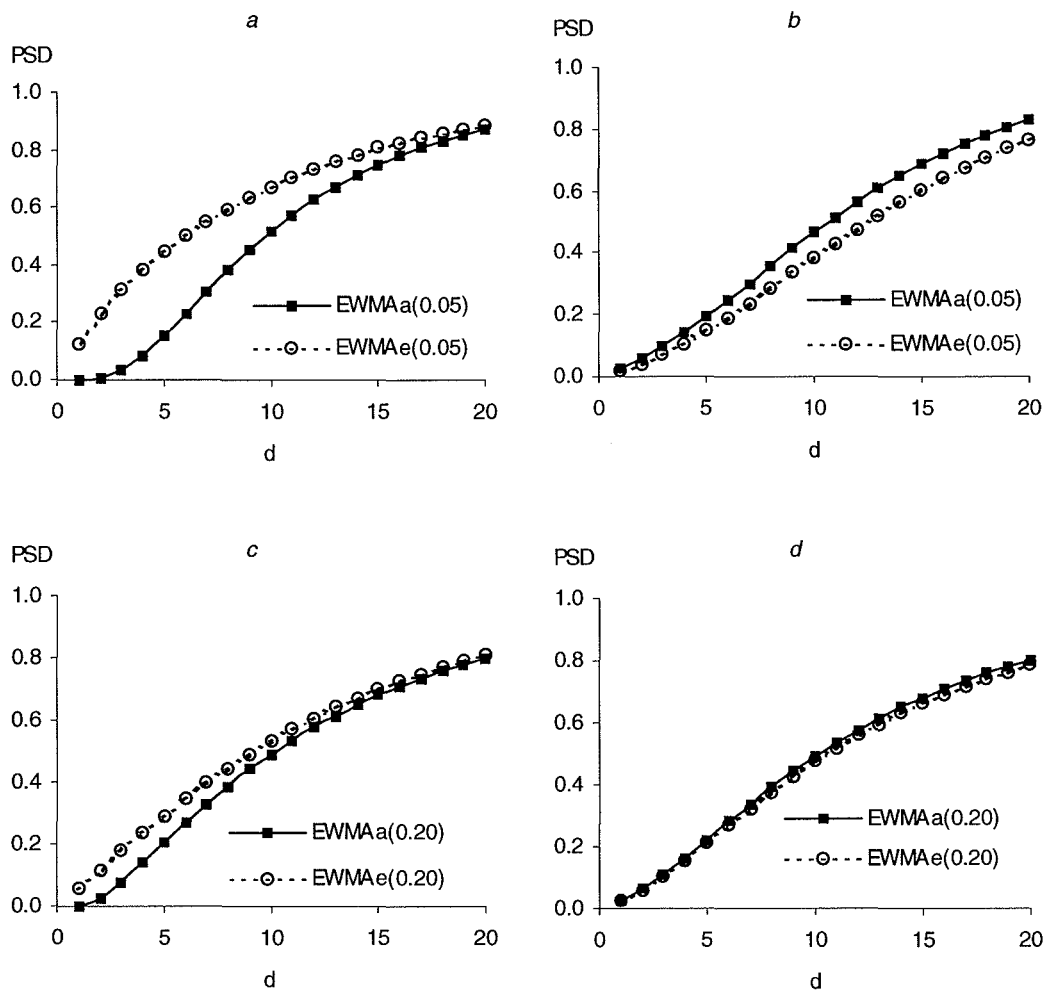


FIG 9:

PSD as a function of d for a fixed value of τ for $\mu = 0.5$ when $P(t_A < \tau) = 0.4877$ for $\nu = 0.01$.

a. $\tau = 1$, b. $\tau = 20$,

c. $\tau = 1$, d. $\tau = 20$.

For the case in Figure 9, the EWMAa is preferable to the EWMAe if the change occurs at a later time point. However, if the change occurs at the first time point the relationship is reversed.

4.4 The confidence to put in an alarm in the one-sided case

One factor, which is often neglected in the evaluation of surveillance methods, is most important when applying the method, namely what to do if an alarm is triggered. The answer to this question depends of course on the application considered. However, when constructing a

method of surveillance, this should be kept in mind. Of major importance is what degree of belief to put in an alarm. In Frisén (1992), the predictive value, $PV(t) = P(C(t) | t_A = t)$ of an alarm was suggested as an evaluation criterion. The motivation is that an alarm with low predictive value should not cause the same actions as one with high predictive value. For the coordination of the actions to follow an alarm, it is preferable if the predictive value is approximately constant over time.

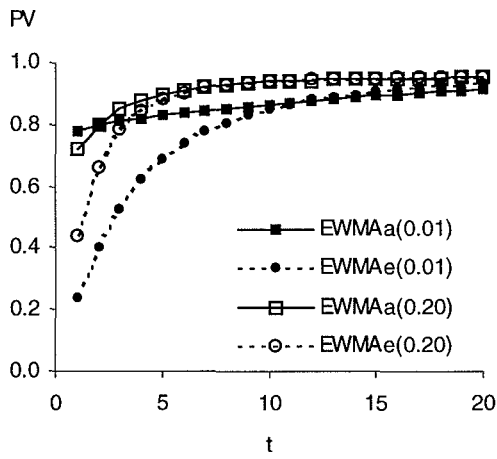


FIG 10:

PV as a function of time for $v = 0.10$, $\mu = 1$, when $ARL^0 = 100$.

One effect of the short delay times for early changes for the EWMAe method is a low predictive value of an alarm at early time points, especially for small values of λ , as illustrated in Figure 10. The EWMAa method has more attractive predictive value features for this case. When $\lambda = 0.20$, the predictive value is not far from constant.

5 TWO-SIDED CASE AND BARRIERS

In this section we will first focus on the different optimal values of λ to minimize the ARL^1 for a fixed value of ARL^0 between the one- and two-sided versions. Secondly, we will consider the minimax properties and also include the one-sided version with a barrier for comparison.

The two-sided version is used in order to detect both positive and negative shifts. Here we study the case of symmetrical alarm limits around 0 using $|Z_s|$ as the alarm statistic. The most obvious difference from the one-sided case is that the value of the alarm statistic at each time point

(conditioned on no alarm) is bounded downwards by the lower alarm limit. This is also the case when using a barrier for the one-sided version. In the usual one-sided version no lower boundary exists. The one-sided version with a barrier to detect positive shifts uses an alarm statistic of the form $\max(b, Z_s)$, where b constitutes the barrier or the lower bound of the alarm statistic. Barriers have important consequences, as will be explored below.

In simulations, we will consider fixed values of ARL^0 of 50 and 100. The constant, L , in the alarm limit will be larger when using a barrier or two-sided alarm limits than for the one-sided version for the same value of λ . However, the difference depends on the value of λ and also on the kind of alarm limits used. The extent to which L is altered will impose a difference in the appearance between the one-sided, the one-sided with a barrier and the two-sided versions of the methods. Worth noting is the agreement of the values of L in this study with those given in Crowder (1989) for the two-sided version of the EWMAa method and thus, these simulations confirm the values given there.

5.1 Differences in the in control features

In Figure 11, the in control run length distributions are presented both for the one- and two-sided case when $ARL^0=50$.

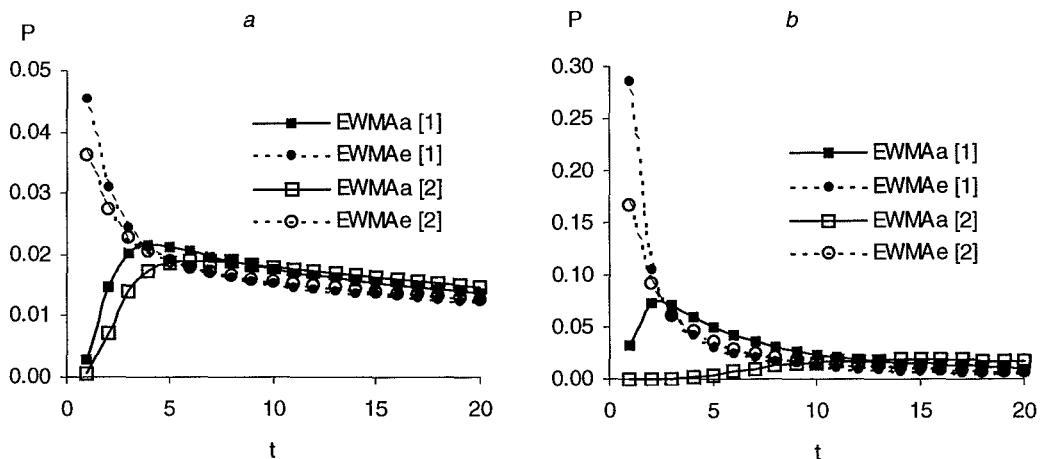


FIG 11:

In control run length distribution when $ARL^0=50$. The value in [] indicates if it is the one- or two-sided version considered.

a. $\lambda = 0.20$, b. $\lambda = 0.01$.

The common feature is that the probability of an early false alarm for the two-sided case is lower than for the corresponding one-sided case. The error spending has thus been shifted towards later time points, due to a change in the dependency structure between successive decisions.

A difference between the EWMAa and the EWMAe method is the mode of the in control run length distribution. For the EWMAe method, the mode is 1 for both the one- and two-sided case. For the EWMAa method, the mode is larger for the two-sided case, especially for small values of λ . This indicates a change also in the ability to detect a true change for the EWMAa method with small values of λ .

5.2 *The differences in detecting a true change*

We now focus on the detection of a true change and explore the differences between the one-sided, the one-sided with a barrier and the two-sided versions of the EWMA methods.

5.2.1 *A change at the first time-point*

When we are in a surveillance situation only considering the ARL^1 , all we have to decide is whether all observations are from the in control or out of control distribution. For the one-sided case, with fixed value of the ARL^0 , this implied equal weight to all observations in order to minimize the ARL^1 . We can compare this situation with a hypothesis test, using a fixed sample, and a sequentially hypothesis testing situation. Both for the hypothesis test and the sequential hypothesis test we want to decide which of the two possible distributions all sampled observations come from.

In a one-sided hypothesis test situation with a fixed sample, equal weight will also be given to all observations for the optimal method. In that case, optimality is usually defined as maximal power for a fixed significance level, as in the Neyman-Pearson Lemma. For a one-sided test with specified means and known variances (simple null and alternative hypotheses) the resulting test statistic is the mean of the observations. This is also the test statistic of the uniformly most powerful test in the case of a composite alternative hypothesis $\mu > \mu_0$. For the two-sided case though, the situation is not that simple. No uniformly most powerful test exists.

In a sequential test situation optimality is often defined in terms of minimal expected sample size, both under H_0 and H_1 , among all tests having no larger error probabilities. In the case of simple null and alternative hypothesis, the resulting test is the SPRT, sequential probability ratio

test. In the case of two normal distributions with specified mean and known variance, the SPRT also results in an optimal test statistic that is the mean of the sampled observations. In this case no uniformly most powerful test exists in the two-sided case.

Therefore, the result in the one-sided case that the value of λ , that minimizes the expected number of sampled units, gives equal weight to all observations in the alarm statistic, independent of the ARL^0 for both the EWMAa and the EWMAe method, should not be surprising.

Many of the previous studies of EWMA methods have been determined to minimize the ARL^1 for a fixed value of ARL^0 for the two-sided case. In Figure 12, values of the ARL^1 are given for a fixed value of ARL^0 . The simulations support the results by numerical approximations in Lucas and Saccucci (1990), Crowder (1989) and Srivastava and Wu (1997) concerning the two-sided EWMAa method with respect to the optimal weighting, λ , of the observations. However, the values presented here suggest that the approximations used in the previous papers overestimate the value of the optimal ARL^1 slightly.

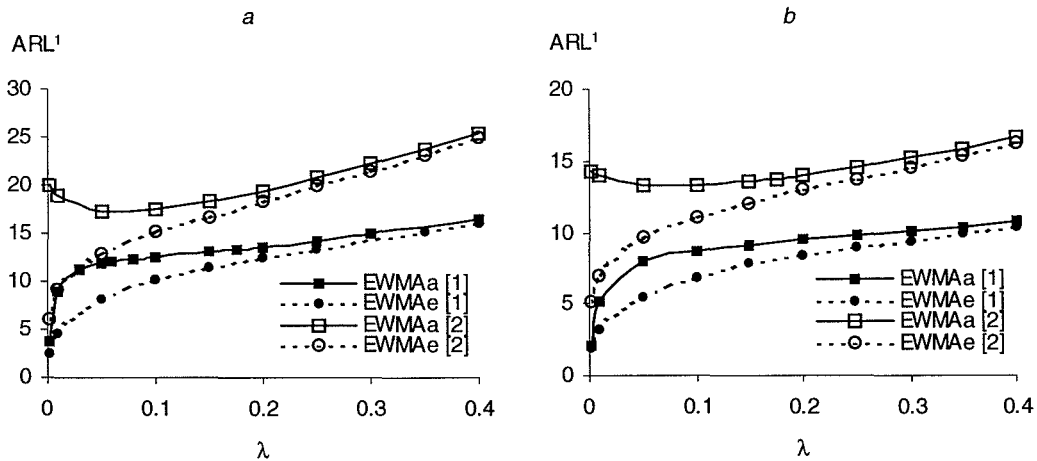


FIG 12: ARL^1 as a function of λ for $\mu = 0.5$.

a. $ARL^0 = 100$. b. $ARL^0 = 50$.

An interesting feature is that the optimal value of λ differs considerably between the EWMAa and EWMAe method for the two-sided case, which was not the case in the one-sided situation. For the EWMAe method the optimal λ still implies equal weight to all observations in the two-sided case. This is no longer the case for the EWMAa method. The difference between the one-sided and two-sided EWMAa is the result of the different error spending as a result of the lower boundary that the two-sided alarm limits imply, which could be seen also in Figure 11 of the in control run

length distributions. However, using a value of λ not equal to zero to distinguish between the in control and out of control distribution, when all observations come from either of them, violates fundamental inference principles.

5.2.2 Detection of changes at later time-points

The use of the lower bound for the alarm statistic will also affect the conditional expected delay of the methods. Here, we focus on the small values of λ and in Figure 13 results are given for $\lambda = 0.01$. For the one-sided case, CED is increasing with time both for the EWMAa and the EWMAe method. For the two-sided case, CED is approximately constant but slightly decreasing over time for the EWMAa while for the EWMAe, CED is increasing with time also for the two-sided case. For both methods, the dependency on time for CED is similar for the two-sided version and the one-sided version with a barrier. Both these versions have a bounded alarm statistic.

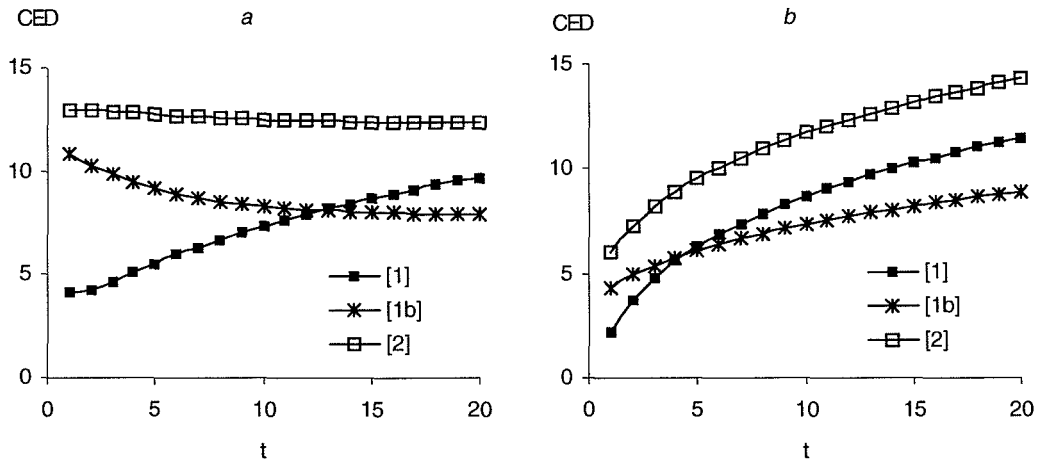


FIG 13: CED as a function of time for $\mu = 0.5$ when $ARL^0 = 50$. The barrier, b , is set to zero in the figure.

a. EWMAa(0.01), b. EWMAe(0.01).

The lower boundary of the alarm statistic for the one-sided case with a barrier and for the two-sided case will also affect the maximum value of CED of the methods. Figure 14a illustrates the maximum value of the CED for the one-sided and two-sided cases when $ARL^0 = 50$ and $\mu = 0.5$. In the case when $\lambda \leq 0.01$ for the one-sided versions, as well as for the two-sided EWMAe (where $CED(t)$ is still increasing with t when $t = 100$), the values of $CED(100)$ are presented. An interesting difference can be seen between the EWMAe and the EWMAa method. For the EWMAa method the two-sided version has considerably lower maximal CED for small values of

λ than the one-sided version. For the EWMAe method this is not true. For the two-sided EWMAa method, the maximal CED is attained at the first time point for all values of λ (see also Figure 13a). In Figure 14b the same situation is illustrated for the EWMAa method with different values of the barrier. Note that the one-sided version is equal to a barrier at $-\infty$. As indicated in Figure 14b, different values of the barrier will be preferred for different values of λ .

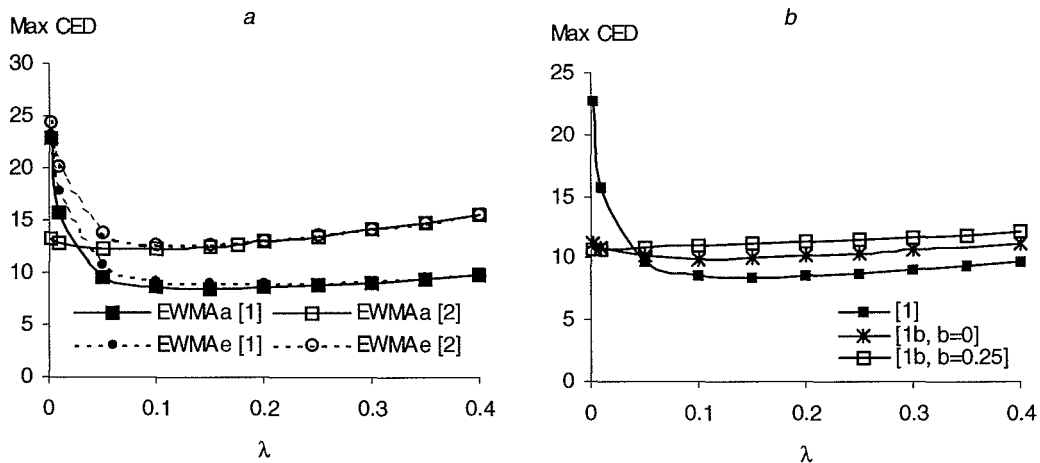


FIG 14: Maximal CED as a function of λ for $\mu = 0.5$ when $ARL^0 = 50$.

6 CONCLUDING REMARKS

The surveillance of a random process to detect a shift in the process has wide-spread application possibilities. To be able to make correct decisions about the state of the process at each time point the help of a properly designed surveillance system is needed. In this paper we have studied different EWMA methods. The focus has been on the effect of different types of alarm limits.

To the EWMA statistic various forms of alarm limits have been suggested in the literature. The most common ones are the EWMAa and the EWMAe. However, the comparative studies between these have only considered shifts at the first time point. In that case, the EWMAe method is preferable to the EWMAa method with respect to the average delay time. However, the predictive value of an early alarm is low for the EWMAe method. When considering also shifts at later time points, as is the natural choice in most applied situations, the picture is changed. In that case the EWMAa method will perform better, both with respect to the conditional expected delay, the expected delay and the probability of successful detection. This shows the importance to consider the performances of methods for shifts at different time points in every surveillance situation.

When comparing the detection ability of surveillance methods a common way is to fix the ARL^0 . Another approach is to compare methods for a fixed probability of a false alarm. Here, it is shown that when comparing the EWMAa and the EWMAe method with respect to the delay time of detecting a true change measured by the expected delay, this is a crucial choice. Choosing to fix the ARL^0 will favour the EWMAe method. If instead the probability of a false alarm is fixed, the EWMAa method performs relatively better.

There are several important differences between the one- and the two-sided versions of the methods. The most striking is the optimal value of λ for minimizing the ARL^1 for a fixed value of ARL^0 . For the one-sided case, λ should approach zero. This is true both for the EWMAa and the EWMAe method. However, this is not the case for the two-sided version of the EWMAa method where the optimal value of λ is larger, thus confirming results in Crowder (1989), Lucas and Saccucci (1990) and Srivastava and Wu (1997). This is somewhat surprising since the minimization of ARL^1 means minimizing the number of observations needed to distinguish between two possible distributions for all observations. To use different weights for different observations violates fundamental inference principles.

Another important difference concerns minimax-properties with respect to the delay time as a function of the time point of the shift. For the two-sided EWMAa method, the alarm limits will act in the same way as when using a one-sided version with barriers, and considerably alter the minimax properties of the method. This is especially apparent for small values of λ . The same pattern is not the case for the EWMAe method.

ACKNOWLEDGEMENT

The author wishes to thank Marianne Friséen for valuable support and comments during the work with this paper. Thanks also to Eva Andersson for valuable comments.

BIBLIOGRAPHY

- Borror, C.M., Montgomery, D.C., and Runger, G.C. (1999). "Robustness of the EWMA control chart to non-normality," *Journal of Quality Technology*, 31, 309-316.
- Box, G.E.P., Jenkins, G.M., and MacGregor, J.F. (1974). "Some recent Advances in Forecasting and Control," *Applied Statistics*, 23, 158-179.
- Champ, C.W. and Rigdon, S.E. (1991). "A Comparison of the Markov-Chain and the Integral-Equation Approaches For Evaluating the Run Length Distribution of Quality-Control Charts," *Communications in Statistics. Simulations and Computations*, 20, 191-204.
- Chan, L.K. and Zhang, J. (2000). "Some issues in the design of EWMA charts," *Communications in Statistics. Simulations and Computations*, 29, 207-217.
- Chandrasekaran, S., English, J.R., and Disney, R.L. (1995). "Modeling and Analysis of Ewma Control Schemes With Variance- Adjusted Control Limits," *IIE Transactions*, 27, 282-290.
- Chang, T.C. and Gan, F.F. (1994). "Optimal Designs of One-Sided Ewma Charts For Monitoring a Process Variance," *Journal of Statistical Computation and Simulation*, 49, 33-48.
- Crowder, S.V. (1987). "A simple method for studying run-length distributions of exponentially weighted moving average charts," *Technometrics*, 29, 401-407.
- Crowder, S.V. (1989). "Design of Exponentially Weighted Moving Average Schemes," *Journal of Quality Technology*, 21, 155-162.
- Crowder, S.V. and Hamilton, M.D. (1992). "An EWMA for monitoring a process standard-deviation," *Journal of Quality Technology*, 24, 12-21.
- Domangue, R. and Patch, S.C. (1991). "Some omnibus exponentially weighted moving average statistical process monitoring schemes.," *Technometrics*, 33, 299-313.
- Frisén, M. (1992). "Evaluations of Methods for Statistical Surveillance," *Statistics in Medicine*, 11, 1489-1502.
- Frisén, M. (1999). "Characterization of methods for surveillance by optimality," Research Report 1999:13 (revised 2000). Department of Statistics, Göteborg University
- Frisén, M. and de Maré, J. (1991). "Optimal Surveillance," *Biometrika*, 78, 271-80.
- Frisén, M. and Sonesson, C. (2001). "Optimal Surveillance with EWMA," *Manuscript*.
- Frisén, M. and Wessman, P. (1999). "Evaluations of likelihood ratio methods for surveillance. Differences and robustness.," *Communications in Statistics. Simulations and Computations*, 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. (1995). "Joint Monitoring of Process Mean and Variance Using Exponentially Weighted Moving Average Control Charts," *Technometrics*, 37, 446-453.
- Gan, F.F. (1998). "Designs of one- and two-sided exponential EWMA charts," *Journal of Quality Technology*, , 55-69.
- Järpe, E. (2000). On univariate and spatial surveillance. Unpublished PhD thesis. Department of Statistics, Göteborg university.
- Lorden, G. (1971). "Procedures for reacting to a change in distribution," *Annals of Mathematical Statistics*, 42, 1897-1908.
- Lu, C.W. and Reynolds, M.R. (1999). "EWMA control charts for monitoring the mean of autocorrelated processes," *Journal of Quality Technology*, 31, 166-188.
- Lucas, J.M. and Saccucci, M.S. (1990). "Exponentially weighted moving average control schemes: properties and enhancements," *Technometrics*, 32, 1-12.
- Morais, M.C. and Pacheco, A. (1998). "Two stochastic properties of one-sided exponentially weighted moving average control charts," *Communications in Statistics. Simulations and Computations*, 27, 937-952.
- Morais, M.C. and Pacheco, A. (2000). "On the performance of combined EWMA schemes for μ and σ : A Markovian approach," *Communications in Statistics. Simulations and Computations*, 29, 153-174.
- Moustakides, G.V. (1986). "Optimal stopping times for detecting changes in distributions," *Annals of Statistics*, , 1379-87.
- Pollak, M. (1985). "Optimal detection of a change in distribution," *Annals of Mathematical Statistics*, 13, 206-227.
- Roberts, S.W. (1959). "Control Chart Tests Based on Geometric Moving Averages," *Technometrics*, 1, 239-250.
- Robinson, P.B. and Ho, T.Y. (1978). "Average Run Lengths of Geometric Moving Average Charts by Numerical Methods," *Technometrics*, 20, 85-93.
- Runger, G.C. and Prabhu, S.S. (1996). "A Markov chain model for the multivariate exponentially weighted moving averages control chart," *Journal of the American Statistical Association*, 91, 1701-1706.
- Schöne, A., Schmid, W., and Knoth, S. (1999). "On the run length of the EWMA scheme: a monotonicity result for normal variables," *Journal of Statistical Planning and Inference*, 79, 289-297.
- Sonesson, C. and Bock, D. (2001). "Statistical issues in public health monitoring- A review and discussion," Research Report 2001:2. Department of Statistics, Göteborg University
- Srivastava, M.S. and Wu, Y. (1993). "Comparison of EWMA, CUSUM and Shiriyayev-Roberts Procedures for Detecting a Shift in the Mean.," *Annals of Statistics*, 21.

- Srivastava, M.S. and Wu, Y. (1997). "Evaluation of optimum weights and average run lengths in EWMA control schemes," *Communications in Statistics. Theory and Methods*, 26, 1253 - 1267.
- Steiner, S.H. (1999). "EWMA control charts with time-varying control limits and fast initial response," *Journal of Quality Technology*, 31, 75-86.
- Tsui, K.L. and Woodall, W.H. (1993). "Multivariate Control Charts Based on Loss Functions," *Sequential Analysis*, 12.

Statistical Issues in Public Health Monitoring- A Review and Discussion

By CHRISTIAN SONESSON and DAVID BOCK

Department of Statistics, Göteborg University, Sweden

SUMMARY

A review of methods, suggested in the literature, for sequential detection of changes in public health surveillance data is presented. Many authors have noticed the need for prospective methods and there has been an increased interest in both the statistical as well as epidemiological literature on this type of problem in the recent years. However, most of the vast literature in public health monitoring deals with retrospective methods. This is especially apparent dealing with spatial methods. Evaluations with respect to the statistical properties of special interest for on-line surveillance are rare. The special aspects of prospective statistical surveillance as well as different ways of evaluating such methods are described. Attention is given to methods including only the time domain as well as methods for detection where observations have a spatial structure. In the case of surveillance of a change in a Poisson process the likelihood ratio method and the Shiryaev-Roberts method are derived.

Key Words: DETECTION; EXPECTED DELAY; INCIDENCE RATE; MONITORING;
PUBLIC HEALTH SURVEILLANCE; SEQUENTIAL METHODS; SPATIAL CLUSTER

1 INTRODUCTION

An important issue in public health is the timely detection and prevention of various types of adverse health events. An example of this is an increased birth rate of babies with congenital malformations. This was especially apparent during the thalidomide tragedy in the early 1960's. An increased incidence rate of diseases, such as asthma or influenza is another example. Other examples include, the increase in bacterial resistance to antimicrobial agents, the spatial clustering of various forms of cancer and different side effects of drugs newly released on the market. In all of these examples, quick detection and prevention is beneficial both at an individual level as well as to society, for example in terms of reduced medical expenditures. Public health surveillance is defined as the ongoing, systematic collection, analysis, and interpretation of out-come specific data essential to the planning, implementation and evaluation of public health programmes, closely integrated with the timely dissemination of these data to those responsible for prevention and control (Thacker and Berkelman, 1988). The need for this type of systems is reflected in the vast and diverse literature of the subject. For example, Blindauer et al. (1999) discuss the need for a nationwide surveillance system for the prevention and control of pesticide-related illness and injury. The risk for adverse health outcomes related to chemical exposures is discussed in Hertz-Picciotto (1996) where the use of an environmental health surveillance system is suggested. Thacker et al. (1996) propose a framework to enhance the practice of surveillance in the United States and discusses current and future surveillance needs.

To be able to control various adverse health events, large amounts of data are collected in various nationwide public health programs such as the National Notifiable Diseases Surveillance System (NNDSS) in the United States controlled by the Centers for Disease Control and Prevention (CDC). In this case 52 different diseases (as of 1 January 1999, Centers for Disease Control and Prevention, 1998) are tracked and data are reported weekly both at state and national level. In England and Wales, the Communicable Disease Surveillance Center (CDSC) and the Public Health Laboratory Service (PHLS) handle these issues. In Hannoun and Tumova (2000), a survey of influenza surveillance systems in 24 European countries is reported. An example is the Groupe Régional d'Observation de la Grippe (GROG) in France, described in Hannoun et al. (1989). Salmonella is also under surveillance in many countries including France and the National Salmonella Reference Centre (NSRC) at the Pasteur Institute in Paris. Another example is bacterial food borne infections. In the United States, collaboration between CDC, Food and Drug Administration (FDA) and the US Department of Agriculture (USDA) has led to the Foodborne Diseases Active Surveillance Network known as FoodNet (Stephenson, 1997). Another type of surveillance system concerns with the safety of marketed drugs. Different type of regulations controls the reporting of drug-related adverse events.

Two examples of this is the Guideline for post marketing reporting of adverse drug experiences by the FDA in the United States (Food and Drug Administration, 1992) and the SAMM Guidelines by the Medicines Control Agency in the UK (Medicines Control Agency, 1993).

Also international networks of centers are in use. An example of this is the World Health Organization (WHO) network of influenza surveillance, FluNet. Other examples include European collaboration of influenza surveillance described in Fleming and Cohen (1996). The International Clearinghouse for Birth Defects Monitoring Systems (ICBDMS) (Erickson, 1991) and EUROCAT are two networks for surveillance of birth defects. For salmonella surveillance an example is the Enter-Net network founded by the European Union under the BIOMED 2 programme. These are only a few examples of the various systems in use today. Further reading can be found in Flahault et al. (1998). In Thacker and Berkelman (1988), a general review of the history and development of public health surveillance in the United States can be found.

The total amount of data collected in these systems is enormous. The data collected can be handled in different ways in order to use the provided information. Common to all applications mentioned and to all public health surveillance systems is that a decision of whether to take preventive actions or not has to be made sequentially based on the data collected so far. From a statistical point of view, this is a much more complicated situation than in a fixed sample situation. For example, traditional hypothesis testing cannot be used. Instead sequential methods such as statistical surveillance should be used.

Different definitions and use of the term surveillance exist in different types of literature. In most of the literature it is not necessary to declare that the term surveillance is used for a prospective situation where observations are gathered sequentially. This is in opposite to a retrospective or fixed sample situation where observations are not accumulating over time. For clarification, by statistical surveillance we mean the prospective or online observation of a stochastic process $X = \{X(t); t = 1, 2, \dots\}$ with the aim of detecting an important change in the process at an unknown time-point τ , as quickly as possible. Much of the research on statistical surveillance was originally done with focus on applications in industrial production control. How commonly these methods are used in practice in public health surveillance systems today can be questioned. For example, Hilsenbeck (1990) reported that none of the examined cancer registers in North America used any statistical control procedure. Instead some informal process control was used. However, the usefulness of statistical surveillance methods also in public health related settings are reported in many papers.

The context of public health surveillance implies specific problems not generally present in the case of an industrial production control. Stroup et al. (1993) notes the problems of seasonal effects and reporting delays in the National Notifiable Diseases Surveillance System. Thacker and Berkelman (1988) discuss problems of incomplete or inaccurate reporting. In Lui and Rudy (1989) and Hillson et al. (1998), the problem of how to handle time lags in case reporting is discussed. Farrington et al. (1996) also address these problems and point out the need for a statistical surveillance system with properties suitable for dealing with problems common in epidemiological data such as bias, delay, lack of accuracy and seasonality. In Thacker et al. (1995), the surveillance of chronic diseases and the requirements for the surveillance system are discussed. It is argued that the characteristics of chronic diseases make the surveillance situation in many aspects different from the one of infectious diseases. In Morabia (1996), the question of what to monitor is raised and the author argues that not only the cases of disease, but also rather the risk factors should be monitored. This type of questions is not exclusive for public health surveillance. The problem of seasonality and delays in reporting was discussed in Andersson et al. (2001) in the case of surveillance of economic time-series. The need for leading indicators was discussed in Andersson et al. (2001) for business cycle surveillance and in Royston (1991) and Andersson (2000) for the use of leading indicators in natural family planning. Although these questions are crucial for a successful statistical surveillance method in a public health context, these issues will not be further discussed here. The review is instead focused on the inferential aspects of proposed statistical surveillance methods. We limit the discussion to the methodological and quantitative part of the surveillance problem and exclude further review of the epidemiological discourse.

Many authors have addressed the problem of constructing methods suitable for public health surveillance. The literature of this subject is found both in statistical as well as in epidemiological journals. The purpose of many papers has been the development of an online monitoring system. However, many of the studies have not taken into account the special statistical aspects of prospective surveillance. Instead the problem has been treated in as if fixed sample situation where data are not accumulating over time. These types of papers are not reviewed here. Some reviews of surveillance methods are already available (Barbujani, 1987) and Farrington and Beale, 1998). However, in Barbujani (1987), the focus is narrowed to methods suggested for surveillance birth effects. In Farrington and Beale (1998), much attention is on key problems when using large surveillance databases. Instead we focus on the inferential part of the surveillance problem including methods for evaluating surveillance methods. A notable feature of many of the methods suggested in the literature is the lack of evaluation by other means than in different case studies. The main purpose of this paper is thus to summarize the current position of surveillance methods in public health related settings and to enhance the use of proper evaluation of methods for surveillance.

The paper is organized as follows. First, in Section 2, some general concepts of statistical surveillance are described. Also different ways of evaluating such a system is presented. In Section 3, the problem of detection of an increased incidence rate is discussed and reviewed. The use of the LR method and the Shiryaev- Roberts method is suggested and the alarm criteria are derived when it is assumed that a Poisson process generates data. In Section 4, the problem of detection of a change in a spatial structure is discussed and reviewed. The current situation is summarized and some concluding remarks are given in Section 5.

2 GENERAL CONCEPTS OF STATISTICAL SURVEILLANCE

By statistical surveillance we mean the online monitoring of a stochastic process $X = \{X(t); t = 1, 2, \dots\}$ with the aim of detecting an important change in the process at an unknown time-point τ , as quickly and as accurately as possible. At each time-point, s , we want to discriminate between two states of the monitored system; the in-control and the out-of-control state, here denoted by $D(s)$ and $C(s)$ respectively. To do this we use the accumulated observations $X_s = \{X(t); t \leq s\}$ to form alarm sets, $A(s)$, such that if $X_s \in A(s)$, this is an indication that the process is in state $C(s)$ and an alarm is triggered. Usually this is done by using an alarm function, $p(X_s)$, and a control limit, $g(s)$, where the time of an alarm, t_A , is written as:

$$t_A = \min\{s; p(X_s) > g(s)\}$$

Different types of in- and out-of-control states are used depending on the application. The most frequently studied case is when $D(s) = \{\tau > s\}$ and $C(s) = \{\tau \leq s\}$. The change to be detected also differs depending on the application. Often a change in a parameter in the distribution of X will be of interest. For example, a change in a parameter can correspond to a changed level, a changed variation or possibly a combined change in the level and variation at the same time. Mostly studied in the literature is the step change, where a parameter changes from one constant level to another constant level. Different types of changes are of interest in different applications. Other type of changes includes a gradual, linear change or an exponential increase.

The alarm function together with the alarm limit constitutes a statistical surveillance method that is a method that tells us when to trigger an alarm, based on the accumulated observations. Thacker et al. (1995) used the term 'surveillance system' to describe a system which include a functional capacity for data collection, analysis and dissemination linked to public health

programmes. Here, we concentrate on the statistical issues of how to handle the information in the data collected, while the term surveillance system is used in a broader sense. For the evaluation of a method of surveillance, different types of measures are used to characterize the behavior both when the process is in- and out-of-control. When the process is in-control, all alarms are false alarms. The distribution of the false alarms is often summarized in the average in-control run length, denoted by $ARL^0 = E[t_A | \tau = \infty]$. Another common measure is the probability of a false alarm, $P(t_A < \tau) = \sum_{t=1}^{\infty} P(\tau = t)P(t_A < \tau | \tau = t)$. However this requires an assumption of the distribution of τ , which often is assumed to be geometric. This assumption is suitable when the probability of a shift at each time point conditioned on no shift before is constant for each time-point.

When evaluating the effectiveness of different types of surveillance methods, one has to face a trade off between false alarms and short delay times for true alarms. The way to handle this is usually in the same way as in a hypothesis-testing situation, where the type 1 error is fixed and evaluations of the power is made for various situations. The translation to the surveillance situation has traditionally been to characterize the type 1 error by the ARL^0 . Then different types of methods have been compared for a fix value of the ARL^0 . Another way of characterizing the type 1 error is by the probability of a false alarm.

The measures of evaluation with respect to a true shift can be made in many different ways. In the vast literature on quality control the average out-of-control run length, $ARL^1 = E[t_A | \tau = 1]$ is usually used. This implies that the change occurred at the same time as the surveillance started. This can be useful in a manufacturing process where one expects various start-up problems. However in a public health situation this is in general not an appropriate approach. In this case one should focus on other measures of evaluation, which takes into account also the possibilities of later changes, since the ability of detection depends on the time-point of the change. One such example is the conditional expected delay as a function of the change point τ :

$$CED(t) = E[t_A - \tau | t_A \geq \tau = t]$$

Assuming a distribution of τ , one could also consider the expected delay:

$$ED_{\tau} = \sum_{t=1}^{\infty} P(\tau = t) \cdot CED(t)$$

In some applications only a limited delay-time can be tolerated. An example is the outbreak of an infectious disease, where an epidemic could be prevented if the outbreak is detected within a given time-interval. In this case we can consider the probability of successful detection:

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

If an alarm is triggered various preventive actions should be taken. To be able to choose what actions to take, it is desirable to know how much trust to put in an alarm. Different surveillance methods have different false alarm distributions as a function of time. Therefore, the proportion of false compared to justified alarms at a specific time point will differ between the methods, that is the trust of an alarm will differ between methods. For choosing what actions to take if an alarm is triggered, the predictive value of an alarm can be used:

$$PV(s) = P(C(s) \mid t_A = s)$$

In Chen et al. (1993), the same type of problem was handled and a method for confirming or rejecting alarms was suggested based on data subsequent to an alarm. However, knowledge of the predictive value could solve this problem without the extra data and thus shorten the time for actions. In general a constant predictive value would be desirable since it would imply that the same actions would be taken whenever the alarm is triggered.

These kinds of measures of evaluation concern the on-line features of the surveillance method. In CDC's guidelines for evaluating surveillance systems (Centers for Disease Control and Prevention, 1988) the timeliness is mentioned as one important aspect when evaluating a surveillance system. It is stated that the timeliness of the surveillance system should be evaluated in terms of availability of information for disease control. This includes both the delay in reporting as well as the time required for the identification of outbreaks. However, no specific measurements for the timely evaluation are provided in the guidelines. Some measures of evaluation are stated, such as the sensitivity and the predicted value positive. These kinds of measures concerns with the ability of correct classification of cases and requires an external source of correct classifications which can be used to validate the data collected by the system. German (2000) gives a review of the use of such measures. These measurements concerns with the quality of the data collected by the surveillance system and do not address the effectiveness of the system to detect adverse events. Therefore they cannot be used as substitutes for the measures of evaluation suggested above.

Further reading on general statistical surveillance can be found in various papers (Shiryaev, 1963; Shiryaev, 1978; Pollak, 1985; Moustakides, 1986; Frisén and de Maré, 1991; Frisén, 1992; Srivastava and Wu, 1993; Siegmund and Venkatraman; 1995; Lai, 1995; Frisén and Wessman, 1999 and Frisén, 1999).

In the following sections, a review of articles covering the topic of statistical surveillance in a public health context is presented. The intention is to summarize the current situation for online surveillance, thus excluding papers dealing with the problem retrospectively. Often the data collected in public health surveillance is represented by counts of cases for example of a disease. This type of data is less studied in most areas of surveillance, where continuous variables are more common. One example though is the case of monitoring the fractions of non-conforming products in an industrial process. A review of methods suggested in this situation can be found in Woodall (1997).

3 DETECTION OF INCREASED INCIDENCE RATES

One major field of research in environmental epidemiology concerns incidence rates. A vast literature covers the production of maps of incidence rates as well as various retrospective tests (Marshall, 1991; Lawson et al., 1999 and Lawson and Cressie, 2000). The literature is rather sparse when it comes to prospective methods of surveillance.

When constructing a surveillance method for detection of an increased incidence rate, different assumptions concerning the underlying process can be made depending on the setting and the data collected. Often, a Poisson process for the cases of disease is assumed. In the case when this assumption has not been considered appropriate, more complex time dependent processes have been used to model the cases of disease. A critical aspect for the system is also whether the baseline rate of the disease is assumed known or not. Based on these assumptions and the type of available data, different types of methods have been suggested for the surveillance, such as the Poisson CUSUM, the Exponential CUSUM, the Sets method and different window methods, which will be further discussed in the coming sections. However, common to all these situations is the sequential decisions to be made at each time point, which make the inferential situation the same.

FIGURE 1 GOES HERE

3.1 *Detection of a Changed Intensity in a Poisson Process*

If a Poisson process for the cases of an adverse health event is assumed, an increased incidence rate corresponds to an increased intensity of the Poisson process. The possibility of detecting such an increased intensity depends both on the way the process is observed as well as the surveillance method used to monitor the process.

3.1.1 *Using the Time between Events to Study the Poisson Process*

In some cases the intervals between the adverse events have been of focus. These intervals can be measured by either the continuous time between the events, which are exponentially distributed, or by using a discrete time scale measuring the number of acceptable events between adverse events. Both these ways includes no loss of information about the process. The increased intensity would then be recognized as shorter intervals between the adverse events and fewer acceptable events between adverse events respectively.

Using the continuous, exponentially distributed time between adverse events, methods like the Exponential CUSUM and the Exponential EWMA can be used. The CUSUM and EWMA methods are two standard methods in statistical process control. Their names come from the way the alarm statistic is formed. For CUSUM, the alarm statistic is based on the cumulative sum of differences between the observations and their expected values. The alarm statistic of the EWMA method is based on an exponentially weighted moving average of the observations. These methods for exponentially distributed variables have not been used in a public health context, but studies have been made in other areas (Vardeman and Ray, 1985; Gan and Choi, 1994; Gan, 1994 and Gan, 1998).

Within the area of surveillance of congenital malformations, the Sets method was proposed in Chen (1978). It focuses on the lengths of the intervals between successive births with malformed babies, measured by the number of healthy babies born between malformed babies. The lengths of these intervals will be geometrically distributed. The method signals an alarm if n consecutive intervals are shorter than some threshold value. In various papers, the Sets method has been further studied (Chen, 1986; Gallus et al., 1986; Radaelli and Gallus, 1989; Sitter et al., 1990; Gallus et al., 1991; Lie et al., 1991; Arnkeldottir, 1995 and Chen et al., 1997). In Arnkeldottir (1995) evaluation was made with respect to the probability of successful detection and the predictive value. In Wolter (1987) and Radaelli (1992), the Cuscore method was studied. In this method a score is assigned of +1 or -1 to each interval between adverse events depending on whether it is longer or not then some threshold value. The alarm statistic is formed from the cumulative score.

However, this type of reporting of the observations means a direct loss of information and a sub-optimal method can be expected.

3.1.2 Using the Number of Events in Fixed Intervals to Study the Poisson Process

If the number of events is recorded for fixed time intervals, information of the process will be lost and the resulting surveillance method will be sub-optimal for detecting the change in the process as quickly as possible. Therefore, using fixed intervals could be motivated only by practical restrictions of the reporting system. For fixed time intervals a commonly used method is the Poisson CUSUM method. It compares the actual number of events in each time period with the expected number and uses the cumulated sum of deviations to form an alarm statistic. A general review of the Poisson CUSUM method can be found in Lucas (1985). The Poisson CUSUM was early applied to congenital malformations in England and Wales (Hill et al., 1968 and Weatherall and Haskey, 1976). In many papers the method has been used to compare and evaluate the performance of alternative methods, for example the Sets method in the previous section (Barbaujani and Calzolari, 1984; Pollak and Kenett, 1983; Gallus et al., 1986; Chen, 1987 and Radaelli, 1992). In Barbujani (1987), these comparisons are reviewed and further described. A sequential binomial likelihood ratio test of the probability that an infant has Down's syndrome was proposed in Lie et al. (1993). In this case the alarm limits were chosen to yield a certain ARL^0 instead of a certain α -level. Comparison with the Poisson CUSUM method was also made with respect to the ARL^1 . In Radaelli (1992), the Poisson CUSUM was compared with the Cuscore method. As an alternative to the Poisson CUSUM, Rossi et al. (1999) evaluated different normal approximations to a Poisson process, in order to improve the method. Other articles discussing the Poisson CUSUM includes Praus et al. (1993) for the use in post-marketing surveillance of adverse drug reactions and Hutwagner et al. (1997) for the case of Salmonella outbreaks. In Bjerkesal and Bakketeig (1975), an early application of the Poisson Shewart method for the case of congenital malformations in Norway can be found.

3.1.3 Observing the Process in a Moving Window

An approach discussed in a retrospective setting in Stroup et al. (1989) and Stroup et al. (1993) was a window-based method. In this case the number of events in a moving window of fixed length is compared with an expected number based on the previous years. This method was suggested for prospective use in Wharton et al. (1993) using data from the National Notifiable Diseases Surveillance System for a four-month period and in Rigau-Perez et al. (1999) for dengue outbreaks in Puerto Rico. Shore and Quade (1989), proposed the SM-method which is based on a

moving window and compared it with the Poisson CUSUM method. However, window based methods are known to be sub-optimal. For example, if one compares two consecutive moving windows of fixed lengths, the ability of detecting a gradual change is low (Sveréus, 1995). Using moving windows will severely reduce the information about the observed process. If the window is wide it will smooth over possible shifts in the process. If, on the other hand, the windows are narrow, the information lost will be larger since only a small amount of the observations are used at each time point. One way of motivating the use of it would be if the base-line rate of the disease were completely unknown.

There are several examples of window-based methods being used in practice. A window-based method was previously in use by the FDA to detect increased frequencies of adverse events related to drugs. In this case the number of reported adverse events in a “report interval” was compared with those of a “comparison interval” and reported to the FDA (Food and Drug Administration, 1992). Recently this type of reporting was revoked (Food and Drug Administration, 1997) with the motivation that the expedited increased frequency reports had not contributed to the timely identification of safety problems. This might be due to the use of a window-based method for detection. Another example of the use is the detection of increased gamma radiation levels in Sweden where two consecutive 24-hour periods are compared by the Swedish Radiation Protection Institute (Kjelle, 1987).

3.1.4 The Likelihood Ratio-Method for a Poisson Process, an Optimal Surveillance Method

The observation of times between events for the Poisson process is preferred to the observation of number of events in fixed intervals if the situation allows for it. However, for the construction of a surveillance method also the alarm statistic and the alarm limits must be considered. The choice of alarm statistic and alarm limits determines the characteristics of the system. The way to choose these is guided by the desired properties of the system, often expressed in terms of an optimality criterion.

In a public health situation, optimality of a surveillance method is not easily determined, due to the complex epidemiological discourse. In our view, the minimization of the expected delay for a fixed probability of a false alarm is a natural choice. Further discussions of optimal surveillance can be found in Friséen and de Maré (1991), Friséen (1999) and Friséen and Sonesson (2000). Consider the case where we want to distinguish between the states $D(s) = \{ \tau > s \}$ and $C(s) = \{ \tau \leq s \}$ for the case of a shift in the intensity of the process from λ_0 to λ_1 . Then this optimality criterion leads to the likelihood ratio method (Friséen and de Maré, 1991). This method

has been studied in some papers, for the case of a positive shift in a normal distribution (Frisén and Wessman, 1999). The time of an alarm for the likelihood ratio method can be expressed as the first time the posterior probability of a change exceeds a constant:

$$t_A = \min\{s; P(\tau \leq s | X_s = x_s) > K\}$$

An equivalent way is the first time the full likelihood exceeds an alarm limit:

$$\begin{aligned} t_A &= \min\left\{s; \frac{f_{X_s}(x_s | C(s))}{f_{X_s}(x_s | D(s))} > \frac{P(\tau > s)}{P(\tau \leq s)} \cdot \frac{K}{1-K}\right\} \\ &= \min\left\{s; \sum_{t=1}^s \frac{P(\tau = t)}{P(\tau \leq s)} \cdot L(s, t) > \frac{P(\tau > s)}{P(\tau \leq s)} \cdot \frac{K}{1-K}\right\} \end{aligned}$$

where $L(s, t)$ is the conditional likelihood at time s for the case when $\tau = t$.

The limitation of the likelihood ratio method is that it requires knowledge of the distribution of the change-point, τ . Often a geometric distribution has been used for other situations. If the intensity of a shift is low, that is, the parameter in the geometric distribution is close to zero, the Shiryaev-Roberts method can be used as an approximation of the likelihood ratio method. This was demonstrated in Frisé and Wessman (1999) to be a good approximation for intensities up to 0.20 in the case of a change in the mean of a normal distribution. The Shiryaev-Roberts method can also be regarded as one, which use a non-informative prior for the time of change.

The time of an alarm for the Shiryaev-Roberts method is:

$$t_A = \min\left\{s; \sum_{t=1}^s L(s, t) > K\right\}$$

where K is a constant.

The likelihood ratio method and the Shiryaev-Roberts method have been suggested in other situations, and the extension to a positive shift in a Poisson process is straightforward. The construction of these methods can be done both in the case when data is represented by the time between events and when data is represented by the number of events in fixed intervals. In both cases, the likelihood ratio and the Shiryaev-Roberts method will be preferable to the previously suggested methods for these situations in the sense that the expected delay will be shorter for a fixed value of the probability of a false alarm. In the case with exponentially distributed time

intervals denoted by X , the time of an alarm for the likelihood ratio method is, for some constant L :

$$t_A = \min\left\{s; \sum_{t=1}^s P(\tau = t) \cdot \exp\left\{-\lambda_1 + \lambda_0 \sum_{i=t}^s X(i)\right\} \cdot \left(\frac{\lambda_1}{\lambda_0}\right)^{s-t+1} > L \cdot P(\tau > s)\right\}$$

For the Shiryaev-Roberts method an alarm will be given at:

$$t_A = \min\left\{s; \sum_{t=1}^s \exp\left\{-\lambda_1 + \lambda_0 \sum_{i=t}^s X(i)\right\} \cdot \left(\frac{\lambda_1}{\lambda_0}\right)^{s-t+1} > L\right\}$$

In the case where the observed data consists of number of events, X , recorded in fixed intervals of length k , for the likelihood ratio method, an alarm will be given at:

$$t_A = \min\left\{s; \sum_{t=1}^s P(\tau = t) \cdot \exp\{(-\lambda_1 + \lambda_0) \cdot k \cdot (s-t+1)\} \cdot \left(\frac{\lambda_1}{\lambda_0}\right)^{\sum_{i=t}^s X(i)} > L \cdot P(\tau > s)\right\}$$

For the Shiryaev-Roberts method the time of an alarm will be:

$$t_A = \min\left\{s; \exp\{(-\lambda_1 + \lambda_0) \cdot k \cdot (s-t+1)\} \cdot \sum_{t=1}^s \left(\frac{\lambda_1}{\lambda_0}\right)^{\sum_{i=t}^s X(i)} > L\right\}$$

If the counts are recorded for intervals of different length, a slight modification has to be done, but again this is straightforward.

For use in an epidemiological context also other properties of these methods needs to be examined properly. A desirable property, which was demonstrated by Frisé and Wessman (1999) to be fulfilled for the Shiryaev-Roberts method, at least in the case of a normal distribution, is that the predictive value is almost constant as a function of time. This would be particularly useful in an epidemiological context and the investigations to follow an alarm as this implies that an alarm could be interpreted in the same way regardless on whether it is late or early. If that is the case also for a shift in a Poisson process could be expected but remains to be verified.

3.2 Processes with Time Dependencies

If the assumption of a Poisson distribution for the cases of disease cannot be motivated, another approach must be taken. Noting that time series of a number of diseases exhibit time dependence (autocorrelation, seasonality etc) a series of papers have been devoted to model these time series. Properly modelled, deviations from the modelled series can be thought as an indication of a change in the disease pattern. Watier et al. (1991) propose an ARIMA type model based warning system where the alert threshold value is a function of the upper side of the prediction interval. The idea was applied to data for Salmonella in France. Nobre and Stroup (1994) use the forecast errors to calculate a probability index function to detect deviation from past observations applied to data for measles cases reported through the NNDSS. In Farrington et al. (1996) a regression algorithm was developed to assist in detecting outbreaks of infectious diseases reported to the CDSC. A threshold for the number of cases was constructed by using prediction intervals for the modeled base-line rate. Evaluation of the detection probability was made. The timely modeling of diseases was also the focus in Williamson and Hudson (1999), where ARIMA models were used on data for various diseases both on national and state level from the NNDSS. The residuals from one-step-ahead forecasts were suggested for surveillance. In VanBrackle and Williamson (1999) this idea was further investigated and the average run length was investigated applying the Shewart, the moving average method and the EWMA method to these residuals for 4 different types of shifts. Other examples of time series modeling can be found in Healy (1983), Ngo et al. (1996), Simonsen et al. (1997), Quenel and Dab (1998) and Cardinal et al. (1999). Reviews of different inferential approaches to the surveillance of processes with autocorrelation or with regression on time or on other variables are found in Frisé (1999).

Other medical problems include kidney failures with various possible changes studied in Smith and West (1983) in a Bayesian framework. Representing the problem as a state space model, the multiprocess Kalman filter was used to calculate on-line posterior probabilities for the different states. Some discussion of how to construct alarm systems based on these probabilities was included. Further reading can be found in Smith et al. (1983) and Gordon and Smith (1990). In Whittaker and Fruhwirth-Schnatter (1994), the same approach was used for detecting the onset of growth in bacteriological infections. An alarm was triggered if the posterior probability of a change exceeded a fix constant. The use of a Shewart-Cusum chart applied to recursive residuals from a continuous time first-order autoregressive, CAR(1), model, where the parameters of the model was continuously updated using a Kalman filter can be found in Schlain et al. (1992). In this case the method was applied to a tumour biomarker. Another examples of this approach can be found in Schlain et al. (1993) and Stroup and Thacker (1993).

4 DETECTION OF A CHANGE IN A SPATIAL STRUCTURE

In most public health surveillance programs, measurements are made at various locations both in space and in time, not only in time. For example, the cases of disease reported to the CDC through the NNDSS are collected at various places all over the US. The data on birth malformations reported by the ICBDMMS consists of data from 35 different countries as of 1 January 2000. This leads to a multivariate situation, with possible spatial dependence between the locations of observation. To deal with this multivariate situation, methods of multivariate surveillance must be used. A multivariate version of the Sets method, using data of malformations from multiple sources, has been proposed (Chen, 1978 and Chen et al., 1982). In this case, fixed time periods was used contrary to the univariate one which uses the time between events. Here the number and size of terminated sets within the time period is used. In Stroup et al. (1988) the possibility of using multiple time series for detection of excess deaths from pneumonia and influenza was discussed. Here, one-step-ahead forecasts were used.

In many cases the methods used to analyse data from surveillance systems prospectively ignores the spatial structure of the data. All of the surveillance methods discussed so far are examples of this. One of the main purposes of the surveillance systems in use is to detect changes in the data observed. If the spatial structure of the data is ignored, this will lead to insufficient and sub-optimal surveillance methods due to loss of information of the observed process. The spatial component in infectious diseases, such as influenza, is clear. An example is the joint collaboration of different European countries during the winter of 1993-1994 (Fleming and Cohen, 1996) where the epidemic started in Scotland and spread south to the rest of the countries via England and France. A considerable time lag in influenza peaks was evident, which could be used for preventive actions.

In many cases the key issue of the public health surveillance itself includes detection of changes in spatial patterns, not only average changes in the case when the data collected are spatially correlated. An example of this is various forms of clustering of diseases of which the case of child leukaemia has been the topic in many retrospective studies. In Dolk (1999) the role of assessing spatial variation and clustering of birth defects is treated. As in the area of detection of increased incidence rates, the area of cluster detection is dominated by different type of retrospective analysis methods, not designed for surveillance (Knox, 1964; Stone, 1988; Besag and Newell, 1991; Lawson, 1993; Waller and Turnbull, 1993; Waller and Lawson, 1995; Tango, 1995; Kulldorff and Nagarwalla, 1995 and Kulldorff, 1997). Papers on the detection of elevated risk due to possible putative sources include Diggle and Rowlingson (1994) and Lawson et al. (1999). However, in many of these situations there would have been an interest also in studying the development prospectively.

To construct surveillance methods for spatial processes is a complicated problem. In previous sections, which only included the time domain, we considered different assumptions of the observed process and different ways of observing and modelling this process. In the spatial domain the same questions are raised. General theory of statistics for spatial data can be found in Cressie (1993). In Lawson (2001) a discussion of how to generalize various kinds of spatio-temporal models to allow for prospective surveillance is given. In the case of spatial surveillance, a change in a parameter of the distribution of the observations can have a clear spatial interpretation, for example, a stronger tendency of clustering.

When confronted with a problem involving both spatial as well as temporal components, which is the case in surveillance of spatial structures, different approaches can be used. One example is the surveillance, in time, of a purely spatial statistic, which describes the spatial pattern for each time-point. This is the case when using a univariate test statistic designed for a retrospective test and following it through time using some surveillance method. This approach was used in Rogerson (1997), where a modification of the retrospective test suggested in Tango (1995) both for general and focused clustering was used prospectively and sequentially with a CUSUM method. The proposed system was evaluated using the ARL and the median run length. The same approach was used in Rogerson (2000) based on the Knox statistic suggested in Knox (1964).

In Raubertas (1989), the spatial structure of the reporting units is taken into consideration, leading to a multivariate surveillance situation. It is argued by the author that when the incidence of a disease is positively correlated between neighbouring reporting units, the sensitivity of the Poisson CUSUM method may be improved by pooling within neighbourhood observations, using closeness as weights. For each reporting unit a Poisson CUSUM is used. For the whole system an alarm is triggered as soon as any of the individual CUSUM schemes signals an alarm. ARL^0 and ARL^1 are suggested as measures of performance.

Another approach is to focus on the spatial model assumed for the observations and to make a sufficient reduction of the spatial structure at each time-point. In this case no information about the spatial structure will be lost. This approach was used in Järpe (1999) in the case of surveillance of clustering in a spatial log-linear model with a fixed lattice. Here the sufficient reduction resulted in the surveillance of a univariate statistic involving the sufficient spatial components for each time. A complete separation of the spatial and the temporal components was possible. The expected delay of an alarm for a fixed false alarm probability was examined for some examples. In Järpe (2000), a shift process spreading spatially as time increased was considered. Here a likelihood ratio statistic was suggested, including a sufficient reduction of the spatial structure. In this case, though, a complete separation of the spatial and temporal components was not possible due to the

nature of the problem. Different ways of treating the multivariate structure in the spatial surveillance situation was discussed. As an application, the problem of an increased rate of radiation was investigated. Some evaluation and comparison with the system currently in use in Sweden, which is based on a moving window was made. The situation with a spreading shift process would correspond well to the surveillance of influenza, where the disease spread across Europe from Scotland (Fleming and Cohen, 1996).

As pointed out also in Lawson (2001), the possibility of development within this area is bright since there are a number of possible applications of statistical surveillance in a spatial context.

5 DISCUSSION AND CONCLUDING REMARKS

The usefulness of properly designed statistical surveillance methods cannot be exaggerated and many authors point out the need for such a system in various public health settings. Except for several practical issues such as the collection of data and the epidemiological investigations to follow an alarm, a surveillance system also raises a number of statistical challenges. Due to the nature of such a system with respect to the sequential type of decision situation, the common retrospective analysis methods are not useful. Many papers have addressed the problem of on-line surveillance but the mistake of not noting the sequential type of decision situation is quite common. In many of the papers, which deal with the inferential aspects correctly, lack of proper statistical evaluation of the suggested methods is evident. Usually, the only measures considered are the ARL^0 and ARL^1 . However, in public health surveillance the event to be detected is not probable to occur at the same time as the surveillance starts. This means that the ARL^1 is not a suitable measure of evaluation. Instead other types of measures should be used, which takes into account also possible later shifts, since the performance of a surveillance method depends on the time of the change.

When constructing a surveillance method theoretically, often the intention is the fulfillment of some optimality criteria. The minimization of the ARL^1 for a fixed value of ARL^0 is the common criteria. The logical drawbacks of this criterion and the advantages of other ones, such as the minimal expected delay for a fixed value of the probability of a false alarm, are discussed in Frisé (1999) and Frisé and Sonesson (2000). In many applications, including public health surveillance, only one measure of performance is not enough. Therefore one should aim at a complete and thorough evaluation of proposed systems. We suggest using measures such as the conditional expected delay, the expected delay, the probability of successful detection and the predictive value.

In practice, data is collected from many different sources, for example in the National Notifiable Diseases Surveillance System. This means that the observed process is multivariate. When discussing the coordination of disease data in different databases, this is a recognized fact (Levy, 1996 and Thacker et al., 1996). However, the proposed surveillance methods in public health, mainly treats the problem as a univariate one. In that way, the dependence between the different observations is not taken into account, which leads to loss of information. Instead, the surveillance situation should be handled as a multivariate one (Wessman, 1998a; Wessman, 1998b and Järpe, 2000).

It is our hope that research within this area is continued since there remains numerous problems to be solved and prospects for development are bright, which will be of great importance for society. In the case of a Poisson process, the properties of the proposed likelihood ratio method and Shiryaev-Roberts method have to be examined properly.

ACKNOWLEDGEMENTS

The authors thank Marianne Frisé and Andrew Lawson for advice and fruitful discussions during the work of this paper.

REFERENCES

- Andersson, E. (2000) Monitoring cyclical processes using non-parametric statistical surveillance. International Biometric Conference 2000. Berkeley, San Francisco, US.
- Andersson, E., Bock, D., and Frisén, M. (2001) Likelihood based methods for turning point detection. A comparative Monte Carlo study. Manuscript.
- Arnkeldottir, H. (1995) Surveillance of rare events. On evaluations of the Sets method. Research Report. 1995:1. Department of Statistics, Göteborg University.
- Barbaujani, G. and Calzolari, E. (1984) Comparison of two statistical techniques for the surveillance of birth defects through a Monte Carlo simulation. *Statist. Med.*, 3, 239-247.
- Barbaujani, G. (1987) A review of statistical methods for continuous monitoring of malformation frequencies. *European Journal of Epidemiology*, 3, 67-77.
- Besag, J. and Newell, J. (1991) The detection of clusters in rare diseases. *J. R. Statist. Soc. A*, 154, 143-155.
- Bjerkedal, T. and Bakketeig, L. S. (1975) Surveillance of congenital malformations and other conditions of the newborn. *International Journal of Epidemiology*, 4, 31-36.
- Blindauer, K. M., Jackson, R. J., McGeehin, M., Pertowski, C. and Rubin, C. (1999) Environmental pesticide illness and injury - The need for a national surveillance system. *Journal of Environmental Health*, 61, 9-.
- Cardinal, M., Roy, R. and Lambert, J. (1999) On the application of integer-valued time series models for the analysis of disease incidence. *Statist. Med.*, 18, 2025-2039.
- Centers for Disease Control and Prevention. (1988) Guidelines for evaluating surveillance systems. *MMVR*, 37, 5, 1-18.
- Centers for Disease Control and Prevention. (1998) Summary of notifiable diseases, United States 1998. *MMWR*, 47, 53, 1-92.
- Chen, R. (1978) A surveillance system for congenital malformations. *J. Am. Statist. Ass.*, 73, 323-327.
- Chen, R. (1986) Revised values for the parameters of the sets technique for monitoring the incidence rate of a rare disease. *Methods of Information in Medicine*, 25, 47-49.
- Chen, R. (1987) The relative efficiency of the sets and the CUSUM techniques in monitoring the occurrence of a rare event. *Statist. Med.*, 6, 517-525.
- Chen, R., Connelly, R. R. and Mantel, N. (1993) Analyzing post-alarm data in a monitoring system in order to accept or reject the alarm. *Statist. Med.*, 12, 1807-1812.
- Chen, R., Iscovich, J. and Goldbourt, U. (1997) Clustering of leukaemia cases in a city in Israel. *Statist. Med.*, 16, 1873-1887.

- Chen, R., Mantel, N., Connelly, R. R. and Isacson, P. (1982) A monitoring system for chronic diseases. *Methods of Information in Medicine*, 21, 86-90.
- Cressie, N. A. C. (1993) *Statistics for Spatial Data*, Revised Edition. New York: Wiley.
- Diggle, P. J. and Rowlingson, B. S. (1994) A conditional approach to point process modeling of elevated risk. *J. R. Statist. Soc. A*, 157, 433-440.
- Dolk, H. (1999) The role of the assessment of spatial variation and clustering in environmental surveillance of birth defects. *European Journal of Epidemiology*, 15, 839-845.
- Erickson, J. D. (1991) The International Clearinghouse for Birth Defects Monitoring Systems: past, present and future. *International Journal of Risk & Safety in Medicine*, 2, 239-248.
- Farrington, C. P., Andrews, N. J., Beal, A. D. and Catchpole, M. A. (1996) A statistical algorithm for the early detection of outbreaks of infectious disease. *J. R. Statist. Soc. A*, 159, 547-563.
- Farrington, C. P. and Beale, A. D. (1998) The Detection of Outbreaks of Infectious Disease. In *GEOMED '97: Proc. Int. Workshop on Geomedical Systems*, (ed. L. Gierl, A. Cliff, A-J. Valleron, C. P. Farrington, and M. Bull), pp. 97-117. Leipzig: Teubner.
- Flahault, A., Parsons, D. F., Garnerin, P. and Valleron, A-J. (1998) Information systems for surveillance of communicable diseases. In *GEOMED '97: Proc. Int. Workshop on Geomedical Systems*, (ed. L. Gierl, A. Cliff, A-J. Valleron, C. P. Farrington, and M. Bull), pp. 157-170. Leipzig: Teubner.
- Fleming, D. M. and Cohen, J. M. (1996) Experience of european collaboration in influenza surveillance in the winter of 1993-1994. *Journal of Public Health Medicine*, 18, 133-142.
- Food and Drug Administration. (1992) Guideline for post marketing reporting of adverse drug experiences. Federal Register, US.
- Food and Drug Administration. (1997) Postmarketing expedited adverse experience reporting for human drugs and licensed biological products; Increased frequency report-FINAL RULE, Federal Register, US.
- Frisén, M. (1992) Evaluations of methods for statistical surveillance. *Statist. Med.*, 11, 1489-1502.
- Frisén, M. (1999) Characterization of methods for surveillance by optimality. Research Report 1999:13 (revised 2000). Department of Statistics, Göteborg University.
- Frisén, M. and de Maré, J. (1991) Optimal surveillance. *Biometrika*, 78, 271-80.
- Frisén, M. and Sonesson, C. (2000) Optimal Surveillance with EWMA, Research Report 2000:7. Department of Statistics, Göteborg University.

- Frisén, M. and Wessman, P. (1999) Evaluations of likelihood ratio methods for surveillance. Differences and robustness. *Communications in Statistics. Simulations and Computations*, 28, 597-622.
- Gallus, G., Mandelli, C., Marchi, M. and Radaelli, G. (1986) On surveillance methods for congenital malformations. *Statist. Med.*, 5, 565-571.
- Gallus, G., Radaelli, G. and Marchi, M. (1991) Poisson approximation to a negative binomial process in the surveillance of rare health events. *Methods of Information in Medicine*, 30, 206-209.
- Gan, F. F. (1998) Designs of one- and two-sided exponential EWMA charts. *Journal of Quality Technology*, 30, 55-69.
- Gan, F. F. (1994) Design of optimal exponential CUSUM control charts. *Journal of Quality Technology*, 26, 109-124.
- Gan, F. F. and Choi, K. P. (1994) Computing average run length for exponential CUSUM schemes. *Journal of Quality Technology*, 26, 134-139.
- German, R. R. (2000) Sensitivity and predictive value positive measurements for public health surveillance systems. *Epidemiology*, 11, 720-727.
- Gordon, K. and Smith, A. F. M. (1990) Modeling and monitoring biomedical time series. *J. Am. Statist. Ass.*, 85, 328-337.
- Hannoun, C., Dab, W. and Cohen, J. M. (1989) A new influenza surveillance system in France - the Ile-De- France Grog .1. Principles and methodology. *European Journal of Epidemiology*, 5, 285-293.
- Hannoun, C. and Tumova, B. (2000) Survey on influenza laboratory diagnostic and surveillance methods in Europe. *European Journal of Epidemiology*, 16, 217-222.
- Healy, M. J. R. (1983) A simple method for monitoring routine statistics, *The Statistician*, 32, 347-349.
- Hertz-Picciotto, I. (1996) Comment: Toward a coordinated system for the surveillance of environmental health hazards. *American Journal of Public Health*, 86, 638-641.
- Hill, G. B., Spicer, C. C. and Weatherall, J. A. C. (1968) The computer surveillance of congenital malformations. *British Medical Bulletin*, 24, 215-218.
- Hillson, E. M., Reeves, J. H. and McMillan, C. A. (1998) A statistical signaling model for use in surveillance of adverse drug reaction data. *Appl. Statist.*, 25, 23-40.
- Hilsenbeck, S. G. (1990) Quality-control practices in centralized tumor registries in North-America. *Journal of Clinical Epidemiology*, 43, 1201-1212.
- Hutwagner, L. C., Maloney, E. K., Bean, N. H., Slutsker, L. and Martin, S. M. (1997) Using laboratory-based surveillance data for prevention: An algorithm for detecting Salmonella outbreaks. *Emerging Infectious Diseases*, 3, 395-400.

- Järpe, E. (1999) Surveillance of the interaction parameter of the Ising model. *Communications in Statistics. Theory and Methods*, 28, 3009-3027.
- Järpe, E. (2000) Detection of environmental catastrophes. Research Report 2000:6. Department of Statistics, Göteborg University.
- Kjelle, P. E. (1987) Alarm criteria for the fixed gamma radiation monitoring stations. SSI Report 87-07. Swedish Radiation Protection Institute.
- Knox, E. G. (1964) The detection of space-time interactions. *Appl. Statist.*, 13, 25-29.
- Kulldorff, M. (1997) A spatial scan statistic, *Communications in Statistics. Theory and Methods*, 26, 1481-1496.
- Kulldorff, M. and Nagarwalla, N. (1995) Spatial disease clusters - Detection and inference. *Statist. Med.*, 14, 799-810.
- Lai, T. L. (1995) Sequential change point detection in quality control and dynamic systems. *J. R. Statist. Soc. B*, 57, 613-658.
- Lawson, A. B. (1993) On the analysis of mortality events around a prespecified fixed point. *J. R. Statist. Soc. A*, 156, 363-377.
- Lawson, A. B., Böhning, D., Lesaffre, E., Biggeri, A., Viel, J.-F. and Bertollini, R. (1999) *Disease Mapping and Risk Assessment for Public Health*. New York: Wiley.
- Lawson, A. B. (2001) *Statistical Methods in Spatial Epidemiology*. New York: Wiley.
- Lawson, A. B. and Cressie, N. (2000) Spatial statistical methods for environmental epidemiology. In *Handbook of Statistics* (ed. P. K. Sen and C. R. Rao), 18, pp. 357-396. Amsterdam: Elsevier Science B.V.
- Levy, B. S. (1996) Editorial: toward a holistic approach to public health surveillance. *American Journal of Public Health*, 86, 624-625.
- Lie, R. T., Heuch, I. and Irgens, L. M. (1993) A new sequential procedure for surveillance of Down's syndrome. *Statist. Med.*, 12, 13-25.
- Lie, R. T., Vollset, S. T., Botting, B. and Skjaerven, R. (1991) Statistical methods for surveillance of congenital malformations: When do the data indicate a true shift in the risk that an infant is affected by some type of malformation? *International Journal of Risk & Safety in Medicine*, 2, 289-300.
- Lucas, J. M. (1985) Counted data CUSUM's. *Technometrics*, 27, 129-144.
- Lui, K. J. and Rudy, R. K. (1989) An application of a mathematical model to adjust for time lag in case reporting. *Statist. Med.*, 8, 259-262.
- Marshall, R. J. (1991) A review of methods for the statistical analysis of spatial patterns of disease. *J. R. Statist. Soc. A*, 154, 421-441.
- Medicines Control Agency. (1993) *SAMM Guidelines: Guidelines for company sponsored safety assessment of marketed medicines*. London: Medicines Control Agency.

- Morabia, A. (1996) Annotation: From disease surveillance to the surveillance of risk factors. *American Journal of Public Health*, 86, 625-627.
- Moustakides, G. V. (1986) Optimal stopping times for detecting changes in distributions. *Ann. Statist.*, 14, 1379-87.
- Ngo, L., Tager, I. B. and Hadley, D. (1996) Application of exponential smoothing for nosocomial infection surveillance. *American Journal of Epidemiology*, 143, 637-647.
- Nobre, F. F. and Stroup, D. F. (1994) A monitoring system to detect changes in public health surveillance data. *International Journal of Epidemiology*, 23, 408-418.
- Pollak, M. (1985) Optimal detection of a change in distribution. *Ann. Math. Statist.*, 13, 206-227.
- Pollak, M. and Kenett, R. S. (1983) On sequential detection of a shift in the probability of a rare event. *J. Am. Statist. Ass.*, 78, 389-395.
- Praus, M., Schindel, F., Fescharek, R. and Schwarz, S. (1993) Alert systems for post-marketing surveillance of adverse drug reactions. *Statist. Med.*, 12, 2383-2393.
- Quenel, P. and Dab, W. (1998) Influenza A and B epidemic criteria based on time-series analysis of health services surveillance data. *European Journal of Epidemiology*, 14, 275-285.
- Radaelli, G. (1992) Using the Cuscore technique in the surveillance of rare health events. *Appl. Statist.*, 19, 75-81.
- Radaelli, G. and Gallus, G. (1989) On detection of a change in the dynamics of rare health events. *Communications in Statistics. Theory and Methods*, 18, 579-590.
- Raubertas, R. F. (1989) An analysis of disease surveillance data that uses the geographic locations of the reporting units. *Statist. Med.*, 8, 267-271.
- Rigau-Perez, J. G., Millard, P. S., Walker, D. R., Deseda, C. C. and Casta-Velez, A. (1999) A deviation bar chart for detecting dengue outbreaks in Puerto Rico. *American Journal of Public Health*, 89, 374-378.
- Rogerson, P. A. (1997) Surveillance systems for monitoring the development of spatial patterns. *Statist. Med.*, 16, 2081-2093.
- Rogerson, P. A. (2000) Monitoring point patterns for the development of space-time clusters. To appear in *J. R. Statist. Soc.*
- Rossi, G., Lampugnani, L. and Marchi, M. (1999) An approximate CUSUM procedure for surveillance of health events. *Statist. Med.*, 18, 2111-2122.
- Royston, P. (1991) Identifying the fertile phase of the human menstrual cycle. *Statist. Med.*, 10, 221-240.

- Schlain, B. R., Lavin, P. T. and Hayden, C. L. (1992) Using an autoregressive model to detect departures from steady states in unequally spaced tumor biomarker data. *Statist. Med.*, 11, 515-532.
- Schlain, B. R., Lavin, P. T. and Hayden, C. L. (1993) Using autoregressive and random walk models to detect trends and shifts in unequally spaced tumor biomarker data. *Statist. Med.*, 12, 265-279.
- Shiryayev, A. N. (1963) On optimum methods in quickest detection problems. *Theory of Probability and its Applications*, 8, 22-46.
- Shiryayev, A. N. (1978) *Optimal stopping rules*. New York: Springer.
- Shore, D. L. and Quade, D. (1989) A surveillance system based on a short memory scheme. *Statist. Med.*, 8, 311-322.
- Siegmund, D. and Venkatraman, E. S. (1995) Using the generalized likelihood ratio statistic for sequential detection of a change-point. *Ann. Statist.*, 23, 255-271.
- Simonsen, L., Clarke, M. J., Stroup, D. F., Williamson, D., Arden, N. H. and Cox, N. J. (1997) A method for timely assessment of influenza-associated mortality in the United States. *Epidemiology*, 8, 390-395.
- Sitter, R. R., Hanrahan, L. P., Demets, D. and Anderson, H. A. (1990) A monitoring system to detect increased rates of cancer incidence. *American Journal of Epidemiology*, 132, 123-130.
- Smith, A. F. M., and West, M. (1983) Monitoring renal transplants: An application of the multiprocess Kalman filter. *Biometrics*, 39, 867-878.
- Smith, A. F. M., West, M., Gordon, K., Knapp, M. S. and Trimble, M. G. (1983) Monitoring kidney transplant patients. *The Statistician*, 32, 46-54.
- Srivastava, M. S. and Wu, Y. (1993) Comparison of EWMA, CUSUM and Shiryayev-Roberts procedures for detecting a shift in the mean. *Ann. Statist.*, 21, 645-670.
- Stephenson, J. (1997) New approaches for detecting and curtailing foodborne microbial infections. *Jama-Journal of the American Medical Association*, 277, 1337-1340.
- Stone, R. A. (1988) Investigations of excess environmental risks around putative sources; statistical problems and a proposed test. *Statist. Med.*, 7, 649-660.
- Stroup, D. F. and Thacker, S. B. (1993) A Bayesian approach to the detection of aberrations in public health surveillance data. *Epidemiology*, 4, 435-443.
- Stroup, D. F., Thacker, S. B. and Herndon, J. L. (1988) Application of multiple time series analysis to the estimation of pneumonia and influenza mortality by age 1962-1983. *Statist. Med.*, 7, 1045-1059.
- Stroup, D. F., Wharton, M., Kafadar, K. and Dean, A. G. (1993) Evaluation of a method for detecting aberrations in public health surveillance data. *American Journal of Epidemiology*, 137, 373-380.

- Stroup, D. F., Williamson, G. D., Herndon, J. L. and Karon, J. M. (1989) Detection of aberrations in the occurrence of notifiable diseases surveillance data. *Statist. Med.*, 8, 323-329.
- Sveréus, A. (1995) Detection of successive changes. Statistical methods in post marketing surveillance. Research Report 1995:2. Department of Statistics, Göteborg University.
- Tango, T. (1995) A class of tests for detecting "general" and "focused" clustering of rare diseases. *Statist. Med.*, 14, 2323-2334.
- Thacker, S. B. and Berkelman, R. L. (1988) Public health surveillance in the United States. *Epidemiologic Reviews*, 10, 164-190.
- Thacker, S. B., Stroup, D. F., Parrish, R. G. and Anderson, H. A. (1996) Surveillance in environmental public health: Issues, systems, and sources. *American Journal of Public Health*, 86, 633-638.
- Thacker, S. B., Stroup, D. F., Rothenberg, R. B. and Brownson, R. C. (1995) Public health surveillance for chronic conditions: A scientific basis for decisions. *Statist. Med.*, 14, 629-641.
- Waller, L. A. and Lawson, A. B. (1995) The power of focused tests to detect disease clustering. *Statist. Med.*, 14, 2291-2308.
- Waller, L. A. and Turnbull, B. W. (1993) The effects of scale on tests for disease clustering. *Statist. Med.*, 12, 1869-1884.
- VanBrackle, L. and Williamson, G. D. (1999) A study of the average run length characteristics of the National Notifiable Diseases Surveillance System. *Statist. Med.*, 18, 3309-3319.
- Vardeman, S. and Ray, D. (1985) Average run lengths for CUSUM schemes when observations are exponentially distributed. *Technometrics*, 27, 145-150.
- Watier, L., Richardson, S. and Hubert, B. (1991) A time series construction of an alert threshold with application to S-Bovismorbificans in France. *Statist. Med.*, 10, 1493-1509.
- Weatherall, J. A. C. and Haskey, J. C. (1976) Surveillance of malformations. *British Medical Bulletin*, 32, 39-44.
- Wessman, P. (1998a) Some principles for surveillance adopted for multivariate processes with a common change point. *Communications in Statistics. Theory and Methods*, 27, 1143-1161.
- Wessman, P. (1998b) The surveillance of several processes with different change points. Research Report 1999:2. Department of Statistics, Göteborg University.

- Wharton, M., Price, W., Hoesly, F., Woolard, D., White, K., Greene, C. and McNabb, S. (1993) Evaluation of a method for detecting outbreaks of diseases in 6 states. *American Journal of Preventive Medicine*, 9, 45-49.
- Whittaker, J. and Fruhwirth-Schnatter, S. (1994) A dynamic change point model for detecting the onset of growth in bacteriological infections. *Appl. Statist.*, 43, 625-640.
- Williamson, G. D. and Hudson, G. W. (1999) A monitoring system for detecting aberrations in public health surveillance reports. *Statist. Med.*, 18, 3283-3298.
- Wolter, C. (1987) Monitoring intervals between rare events: A cumulative score procedure compared with Rina Chen's Sets technique. *Methods of Information in Medicine*, 26, 215-219.
- Woodall, W. H. (1997) Control charts based on attribute data: Bibliography and review. *Journal of Quality Technology*, 29, 172-183.

Research Report

- | | | |
|--------|---------------------------------------|--|
| 2000:1 | Hatemi-J, A. & Shukur, G.: | Multivariate based causality tests of twin deficits in the US. |
| 2000:2 | Petzold, M.: | Aspects on tolerance limit estimation - some common approaches and flexible modeling. |
| 2000:3 | Andersson, L.: | Statistical test of the existence of a turning point. |
| 2000:4 | Almasri, A.: | Some aspects of wavelet analysis in time series. |
| 2000:5 | Järpe, E.: | Surveillance of spatio-temporal patterns. Change of interaction in an Ising dynamic model. |
| 2000:6 | Järpe, E.: | Detection of environmental catastrophes. |
| 2001:1 | Holgersson, H.E.T.: | On assessing multivariate normality. |
| 2001:2 | Sonesson, C. & Bock, D.: | Statistical issues in public health monitoring – A review and discussion. |
| 2001:3 | Andersson, E.: | Turning point detection using non-parametric statistical surveillance. Evaluation of some influential factors. |
| 2001:4 | Andersson, E. & Bock, D.: | On seasonal filters and monotonicity. |
| 2001:5 | Andersson, E., Bock, D. & Frisén, M.: | Likelihood based methods for detection of turning points in business cycles. A comparative study. |
| 2001:6 | Sonesson, C.: | Evaluations of some exponentially weighted moving average methods. |