A CUSUM procedure for detection of outbreaks in Poisson distributed medical health events

Jonsson, R.

A CUSUM PROCEDURE FOR DETECTION OF OUTBREAKS IN POISSON DISTRIBUTED MEDICAL HEALTH EVENTS

Robert Jonsson

Abstract

CUSUM procedures which are based on standardized statistics are often supposed to have expectation zero and being normally distributed. If these conditions are not satisfied it can have serious consequences on the determination of proper alarming bounds and on the frequency of false alarms. Here a CUSUM method for detecting outbreaks in health events is presented when the latter are Poisson distributed. It is based on a standardized statistic with a bias from zero that can be neglected. The alarming boundaries are determined from the actual distribution of the statistic rather than on normality assumptions. The boundaries are also determined from requirements on the probability of false alarms instead of the common practice to focus on average run lengths (ARLs). The new method is compared with other CUSUM methods in Monte Carlo simulations. It is found that the new method has about the same expected time to first motivated alarm and the same sensitivity. However, the new method has expected times to first false alarm that are 9 % – 90 % longer. The new method is applied to outbreaks of sick-listening and to outbreaks of Chlamydial infection.

Key words: Reference value k, sampling- and calibration periods.

1. Introduction

In health statistics it is important to detect outbreaks of diseases as early as possible. To this end a large number of statistical methods have been developed for continuous monitoring of the incidence of health events. A common characteristic of the methods is that the values of a statistic, being a function of the original observations, are calculated sequentially in time and if a value reaches above a certain alarm boundary an alarm is signaled. The mean level of the original observations before outbreak has been termed baseline, acceptable level or in-control level and the mean level after outbreak has been termed unacceptable or out-of-control. Since the pioneer work by Shewhart[25] many different statistics have been suggested. These have been based on averages, moving averages, cumulative sums and likelihood ratios[4,9,22], just to mention a few examples. A somewhat different approach for detecting deviations from a basic level has been to construct tolerance limits, with or without utilizing the longitudinal structure of data[12,26,28].

One popular sequential method for monitoring health events is the cumulative sum (CUSUM), originally proposed by Page[20]. From the series of original data, (Z_t), a new series (S_t) is obtained from the relation

$$S_t = \max(0, Z_t - k + S_{t-1}), \ S_0 = 0 \tag{1}$$

Here k is a reference value that is used to balance the series (S_t). An alarm is signaled as soon as $S_t > h$, where h is a predetermined alarm boundary. Proper values of k and h has customary been found from requirements on average lengths of times to first false alarm (before outbreak) and to first motivated alarms (after outbreak). Tables for choosing k and h in this way have been published for the case when the Z_t's have a normal distribution[8] or a Poisson distribution[8,16]. In this paper, where the focus is on health events rather than on industrial processes, other criteria for choosing k and h will be given.

A stumbling block when determining k and h is that the mean (acceptable) level before outbreak and the mean (unacceptable) level after outbreak have to be specified. CUSUM was originally designed for controlling industrial processes, in which case the definition of acceptable and unacceptable levels can be based on previous experience and economical considerations. When dealing with data of health events, such definitions may be harder to

make. Two examples of this are given in Section 5 of this paper, one concerns outbreak of sick-listening and another concerns outbreak of Chlamydia. Here it is obviously a delicate task to decide what levels are acceptable and unacceptable. Some guidelines on this topic are given in Section 4, despite of the difficulties involved.

Lorden proved that the CUSUM method is optimal for one-sided tests in all baseline distributions[15]. The latter conclusion was reached by using a criterion that might be called 'worst average detection delay' (cf. p. 281 in (10) for a formal definition). Later Moustakides proved that CUSUM is exactly optimal as judged by Lorden's criterion[17]. However, it is hard to apply these theoretical results since Lorden's criterion refers to situations that rarely happen in practice.

A large number of Monte Carlo simulation studies with Poisson distributed health events have been published where the CUSUM method is compared with other methods, but these do not give any conclusive results. Barbujani and Calzolari compares CUSUM with a method called 'the sets technique' and finds that CUSUM is more sensitive to real increases and less likely to issue false alarms[1]. Hutwagner et al. came to the same conclusion when CUSUM was compared with Shewart- and moving average charts[11]. A bit more complex results can be found in other studies. Chen found that CUSUM is more efficient than 'the sets technique' for high frequencies, but not for low[3]. Han et al. found that CUSUM was better than EWMA for large shifts, but not for small[10]. Perry and Pignatiello concluded that, when CUSUM was compared with EWMA and a Maximum Likelihood estimate of the time change, neither of the methods performed uniformly best when the relative increase of the mean varied between 20 % and 75 %[21]. The perhaps worst example of the performance of CUSUM can be found in a study by Choi et al.[4]. Here CUSUM was compared with six other methods, using five evaluation measures, and it was found that the performance of CUSUM was poor and sometimes beaten by very simple methods such as 'historical limits'. This mix-up of results in simulation studies calls for that more attention should be given to the conditions under which the simulations are performed. E.g. in some of the studies mentioned it is assumed that at least the mean of the baseline is a known constant, and sometimes even that the mean after outbreak is known. In practice parameters have to be estimated from data of health events and this give rise to a certain amount of uncertainty which ultimately affects the choices of proper values of k and h. Some studies use the original Poisson observations, while other use standardized variables, assumed to be normally distributed with mean zero and variance 1.

If the normality assumption does not hold, this can have a large effect on the choices of k and h. A further fact that seems to have been overlooked is that the series in Eq. (1) is not stationary from start. Instead, the mean of S_t increases with t until the mean stabilizes, in the absence of any outbreak. This in turn increases the possibility of having false alarms.

In this paper a CUSUM procedure is suggested that is based on a standardized Poisson variable. Section 2 deals with the problems of making the latter variable unbiased for 0, to find proper lengths of the sampling period when estimating the baseline mean, to study the time until the CUSUM statistic has stabilized and to determine the values of k and h. In Section 3 some measures of the ability to detect outbreaks are discussed. The results from Monte Carlo simulations are presented in Section 4. Here also the performance of the proposed CUSUM statistic is compared with a similar statistic that was suggested earlier by Rossi et al.[24]

2. Proposed test procedure

2.1 Choice of a proper standardized Z-statistic

Let Y be a Poisson variate with constant mean α_0 before outbreak and mean α_1 after outbreak, where $\alpha_1 (> \alpha_0)$ is some increasing function in time. $\hat{\alpha}_0 = \sum_{i=1}^{n} Y_i / n$ is an estimator of α_0 based on independent observations before outbreak and it is assumed that Y and $\hat{\alpha}_0$ are independent. Consider the following standardized Z-statistics that are supposed to have zero means and where large values are supposed to signal an outbreak,

$$Z^{(1)} = \frac{Y - \hat{\alpha}_0}{\sqrt{\hat{\alpha}_0}}, \quad Z^{(2)} = 2\left(\sqrt{Y} - \sqrt{\hat{\alpha}_0}\right), \quad Z^{(3)} = \frac{Z^{(1)} + Z^{(2)}}{2}, \quad Z = \frac{Y - \hat{\alpha}_0 - 1/2n}{\sqrt{\hat{\alpha}_0}} \qquad (2)$$

Here $Z^{(1)}$ is an ordinary standardized Poisson variate but with α_0 replaced by its estimator. Due to this it will not have zero expectation. The bias of $Z^{(1)}$ before outbreak is approximately $\left(2n\sqrt{\alpha_0}\right)^{-1} > 0$ (See Appendix A.). Thus, there is an increased risk of getting over-estimated values of $Z^{(1)}$ and thereby false alarms if n and α_0 are small. $Z^{(2)}$ is a further variate that has been proposed[14]. Nor the latter has zero expectation and the bias before outbreak is approximately $-(1-1/n)\left(4\sqrt{\alpha_0}\right)^{-1} < 0$ (Appendix A). By using $Z^{(2)}$ the risk of getting false alarms would thus be smaller than for $Z^{(1)}$, but the problem is that the negative bias persists after the outbreak (Appendix A), so the chance of having motivated alarms is expected to be reduced. $Z^{(3)}$ has been suggested by Rossi et al[24] (although the α-parameter was treated as a known constant). In the latter paper it was concluded from Monte Carlo simulations that this statistic performed better than $Z^{(1)}$ and $Z^{(2)}$. By averaging two statistics, one with a positive bias and one with a negative bias, one might obtain an improvement. However, in Table 1 it is obvious that none of $Z^{(1)}, Z^{(2)}$ and $Z^{(3)}$ has a bias that can be ignored unless n and α_0 is large. The statistic Z in Eq. (2) is an attempt to reduce the bias of $Z^{(1)}$ by subtracting the estimated bias. As can be seen in Table 1, this trial proves successful and in the sequel the statistic Z will be used.

One purpose of standardizing is to construct a variate that can be considered normally distributed with mean 0 and variance 1 (N(0,1)). It is easily shown that all statistics in Eq. (2) converge in distribution to N(0,1) (cf. Ch. 20.2 and 20.6 in (7)). In such a case one may utilize the nomograms published by Ewan and Kempf[8] for determining optimal alarm bounds. The problem is that normality may not be achieved unless α_0 and n are large. It is beyond the scope of this article to study the approach to normality in detail, but it may be pointed out that among the statistics in Eq. (2) it was $Z^{(3)}$ that turned out to converge most rapidly to N(0,1). With $\alpha_0 = 5$ and $n = 10$ the 95 % and 99 % percentiles in the distribution of $Z^{(3)}$ were 1.73 and 2.56, respectively. These are however far from the corresponding values 1.64 and 2.33 in the N(0,1) distribution. The normality assumption will only be used for comparisons in this paper.

Table 1 Bias (from zero) of the four Z-statistics in (2) for various values on n and α_0. Each figure is the average from 100,000 simulations. The theoretical approximate biases (see Appendix A) are very close to the ones given in the table.

n	α_0	$Z^{(1)}$	$Z^{(2)}$	$Z^{(3)}$	Z
5	5	.048	-.107	-.030	.002
	10	.035	-.066	-.015	.002
	30	.019	-.036	-.008	.001
	100	.010	-.021	-.005	-.000
10	5	.016	-.126	-.055	-.002
	10	.019	-.079	-.036	.002
	30	.011	-.045	-.016	.001
	100	.010	-.021	-.008	.000
30	5	.013	-.121	-.054	.001
	10	.010	-.077	-.035	.001
	30	.004	-.045	-.019	.001
	100	.001	-.027	-.012	-.000
100	5	.001	-.133	-.065	-.001
	10	.003	-.088	-.037	.001
	30	-.001	-.053	-.023	-.001
	100	.004	-.018	-.019	.000

2.2 Choice of reference value k and sampling- and calibration periods

Often one wants to start the surveillance as soon as possible, but there are two stumbling blocks to deal with. First, the sampling period should be long enough so that a reliable estimate of the baseline rate α_0 is obtained. Second, given a reliable estimate of α_0 one

has to consider the fact that the sequence (S_t) defined in eq. (1) is not stationary from start, and needs some time to reach a stationary level.

Consider first the precision of the estimated rate before outbreak. Table 2 shows the 5 % and 95 % percentiles in the distribution of $\hat{\alpha}_0$, which may be used as a crude measure of the probability concentration of the estimates. Although the concentration around the true value steadily increases with increasing sample size, it is seen that little is gained by increasing n from e.g. 30 to 100 especially for larger values of α_0.

Table 2 5 % - 95 % percentiles in the distribution of $\hat{\alpha}_0$ determined from 100,000 simulations.

		n			
		5	10	30	100
α_0	5	3.4 - 6.8	3.9 - 6.2	4.3 - 5.7	4.6 - 5.4
	10	7.8 - 12.4	8.4 - 11.6	9.1 - 11.0	9.5 - 10.5
	30	26 - 34	27 - 33	28 - 32	29 - 31
	100	93 - 107	95 - 105	97 - 103	98 - 102

The standard rule for choosing the reference value k using a normal variate is to put $k = (m_a + m_r)/2$, where m_a and m_r are the expected levels that are acceptable (during base line) and out-of-control (after an outbreak), respectively (cf. p. 372 in (8)). In cases when no obvious out-of-control level can be specified one may study how k behaves as a function of the relative increase of the mean, $RI = \alpha_1 / \alpha_0 - 1$, that one is aiming to detect. The latter is a function of the time that elapsed from the outbreak. From the results in Appendix A it follows that k for the statistic $Z^{(3)}$ is obtained as $k = \left[E\left(Z^{(3)}|R=1\right) + E\left(Z^{(3)}|R>1\right) \right]/2 \approx C_{\alpha_0} \cdot RI$. In the last expression C_{α_0} is a constant that depends on α_0, but the expression is roughly the same for all n between 5 and 100. A similar relation holds for determining the reference value for the statistic Z. Proper values

of k can be found from Table 3. To take an example, it is required to detect outbreaks corresponding to a value of RI of about 0.5 to 1.5 when $\alpha_0 = 10$. Since $k \approx 1.59 \cdot RI$ one should chose a reference value somewhere between 1 and 2. This thumb rule may seem to be a bit naïve and hard to use in practise, but it gives anyhow rough information about the magnitude of k. Below it will be seen there are also other aspects that determine the choice of k.

Table 3 Values of C_{α_0} in the relation $k \approx C_{\alpha_0} \cdot RI$ for some values of α_0. For other values of α_0 linear interpolation may be used.

	$\alpha_0 =$						
	5	10	20	30	40	50	100
k for $Z^{(3)}$	0.98	1.38	1.95	2.39	2.76	3.08	4.35
k for Z	1.13	1.59	2.24	2.74	3.17	3.54	5.00

Given that α_0 has been estimated one can not start the surveillance process immediately because the sequence $(S_t)_{t \geq 1}$ is not stationary (up to 2nd order) from start, but reaches a stationary state after some time. During the non-stationary part the mean of S_t increases with t, as well as the variance. It is obvious that the risk of false alarms is higher if the surveillance process starts during the non-stationary part before outbreak since the increase in S_t may be confused with a real change of the α-parameter. Figure 1 illustrates the dependency on the reference value k in order to reach stationarity. For $k = 1.0$ and 1.1 stationarity seems never to be reached for t less than 20, although the increase of the mean is very small for large t. For $k = 1.3$ stationarity is reached at about $t = 13$ and for $k = 1.5$ at about $t = 6$.

To study the approach to stationarity more in detail for various values of k and α_0, 100,000 simulations were performed to calculate the mean of S_t for $t = 1, ..., 20$ using a sampling period of $n = 10$ and with α_0 being varied from 1 to 200. It was found that

approximate stationarity was found within a calibration period of 20 time units if $\alpha_0 \geq 5$ and $k \geq 1.1$. For α_0 less than 5, k has to be chosen larger. For $\alpha_0 = 4,3,2,1$ k has to be at least 1.1, 1.5, 2.1, 3.1, respectively, in order to reach stationarity within the given calibration period. As a curiosity one can mention that for $\alpha_0 = 1$ and $k = 1.0$ stationarity was not reached within 100 time units.
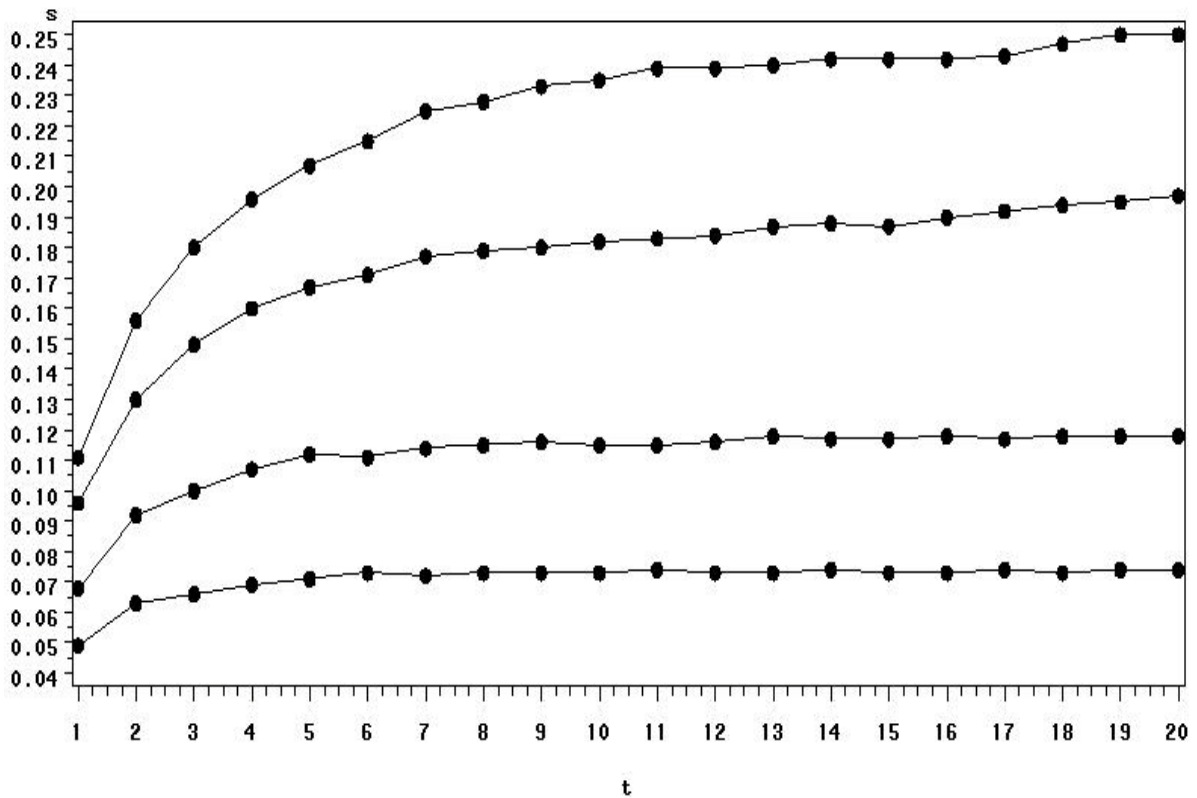

Figure 1 Expectation of S_t for $t = 1 \ldots 20$. The uppermost to the lowest curve shows the expectation for $k = 1.0, 1.1, 1.3, 1.5$ and $\alpha_0 = 10$ ($n = 10$). Each point in the figure is the mean calculated from 100,000 simulations.

2.3 Determination of the alarm boundary *h*

An alarm for an outbreak is given if $S_t > h$ in (1). When studying a change from one ('in control') mean m_a to another ('out of control') mean m_r, k and h have customary been chosen to achieve a specified average run length (ARL) (that is, average time until an

alarm is signaled) under normality assumptions for the Z- observations[8]. In the case considered in this paper there is no well-defined value of m_r and furthermore, the normality assumption seems unrealistic unless α_0 is large. Instead h is chosen so that the probability of a false alarm (FA) at a single time point, say P_{FA}, is at most 5 %, 1 % or 0.5 %. Table 4 shows the values of h determined from simulations that meet those requirements for some values of k and α_0.

Table 4 Values of h such that the probability of a false alarm (that is, the probability that $S_t > h$ before outbreak) at a single time point is 5 %, 1 % and 0.5 %. For given k and α_0 the h values corresponding to the three false alarm probabilities were determined from 100,000 simulations. In each simulation a sampling period of $n = 10$ and a calibration period of 20 were used.

α_0	5 %			1 %			0.5 %		
	$k = 1.1$	$k = 1.3$	$k = 1.5$	$k = 1.1$	$k = 1.3$	$k = 1.5$	$k = 1.1$	$k = 1.3$	$k = 1.5$
5	1.40	0.94	0.61	3.30	2.39	1.90	4.52	3.15	2.48
10	1.28	0.84	0.53	2.98	2.12	1.63	3.94	2.75	1.16
15	1.22	0.79	0.49	2.79	2.00	1.51	3.62	2.57	1.96
20	1.18	0.78	0.47	2.75	1.91	1.46	3.57	2.48	1.87
30	1.14	0.76	0.44	2.52	1.88	1.42	3.33	2.45	1.84
40	1.12	0.74	0.43	2.49	1.81	1.40	3.25	2.35	1.79
50	1.11	0.72	0.41	2.47	1.79	1.36	3.22	2.28	1.70
75	1.09	0.69	0.39	2.44	1.75	1.30	3.20	2.21	1.67
100	1.03	0.67	0.37	2.32	1.62	1.22	3.11	2.06	1.56
150	1.02	0.63	0.36	2.28	1.61	1.20	2.92	2.05	1.55
200	1.02	0.63	0.35	2.28	1.61	1.20	2.91	2.04	1.54

For given k, the h-values decrease somewhat irregularly with increasing α_0 and seem to stabilize at about $\alpha_0 = 200$. In practise α_0 is estimated. The proper h-value corresponding

to the latter estimate can easily be found by linear interpolation. For estimates above 200 the *h*-values at the bottom line can be used. The table does not show *h*-values for α_0 less than 5. The reason for this is that the calibration period now has to be extended (cf. the preceding section) and this in turn leads to a substantial increase of the computer time for the simulations.

2.4 An illustrative example

To illustrate the procedures in Sections 2.2-2.3, consider the following example where a sequence $(Y_i)_{i=1}^{40}$ of Poisson variates is simulated using the means $\alpha_0 = 10$ for $i < 35$ and $\alpha_1 = \alpha_0 \exp\{0.1(i - 35 + 1)\}$ for $i = 35,...,40$, i.e. there is a weak exponential outbreak starting at $i = 35$. The first step is to estimate α_0 from the $n = 10$ first observations giving the estimate $\hat{\alpha}_0 = 9.3$. The second step is to use the transformation Z in (2) to get a new sequence of standardized variates $\{Z_i\}_{i=11}^{40}$. The Z_i's are plotted as dots in Figure 2. Next, the sequence of CUSUM statistics $S_i = \max(0, Z_i - k + S_{i-1})$ is calculated for $i = 11,...,40$, with $S_{10} = 0$ and $k = 1.1$. The CUSUM statistics are plotted as circles in Figure 2. Finally one has to determine the boundary *h* which is the alarm limit for the CUSUM statistics. In this case it is decided to accept a probability of false alarm at a single time point of at most 5 %. From Table 3 $h = 1.296$, obtained from linear interpolation using $\hat{\alpha}_0 = 9.3$. This limit is depicted in Figure 2.

In the figure a typical feature of CUSUM statistics can be seen, compared with the original Z_i's. The former are more conservative in the sense that they have less tendency to react on accidental changes. In Figure 2 one may notice that CUSUM did not alarm at $i = 35$ (the time for the outbreak), but at $i = 36$ and later.
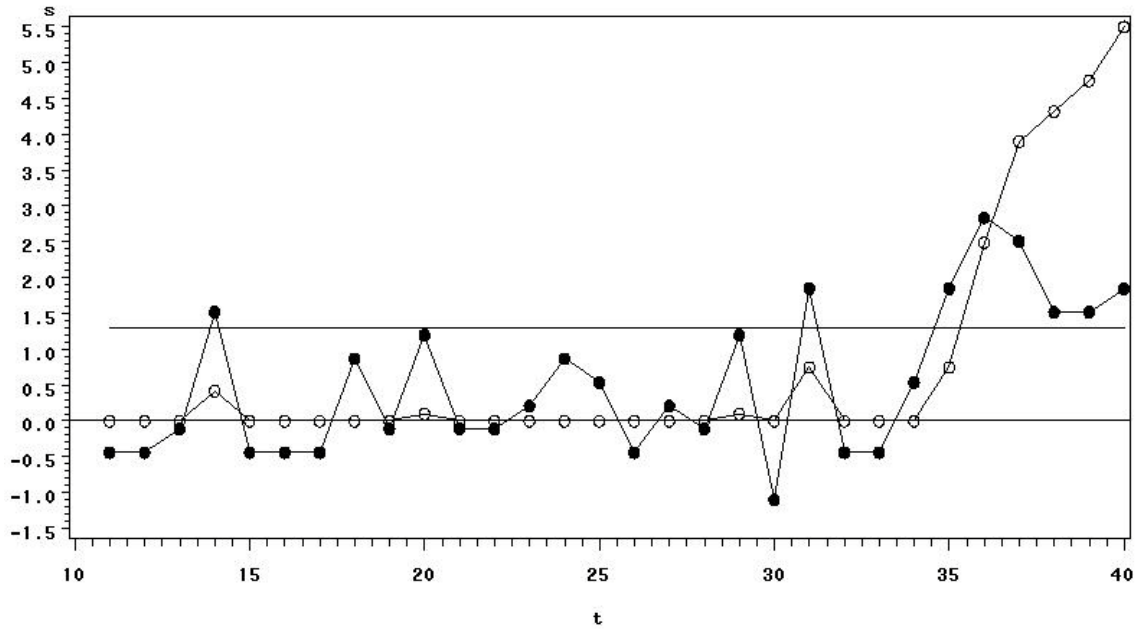
Figure 2 Illustration of the CUSUM statistic when used to detect an outbreak at $t = 35$ (unfilled circles) together with the observed standardized Poisson variates (filled circles). The horizontal line is the alarm limit $h = 1.296$.

3. Measures of ability to detect outbreaks

A large number of evaluation measures have been suggested to study the ability to detect outbreaks. These have mainly been used for comparing the performance of different methods, or in simulation studies to find optimal values of the parameters that are involved in a specific method, in this case k and the probability of a false alarm (FA) at a single time. Here the following aspects are considered: (1) Times-between-FAs, (2) $T_{FA} =$ Times to first FA with expectation E_{FA}, (3) $T_{MA} =$ Times to first motivated alarm, i.e. times to first alarm after outbreak, with expectation E_{MA}, (4) Sensitivity (*Sens*) = Probability of an alarm given that an outbreak has occurred, (5) Specificity (*Spec*) = Probability of no alarm given that no outbreak has occurred, (6) Positive predictive value (*PPV*) = Probability of an outbreak given an alarm, (7) Negative predicted value (*NPV*) = Probability of no outbreak given no alarm.

3.1 Instability of measures based on Times-Between-False Alarms

In a preparatory study the distribution of times-between-FAs was studied in 10 replicates with $\alpha_0 = 10$, each of length 10^6 time units. The probability of a FA at a single time point, P_{FA}, was chosen as 5 % and the reference value was $k = 1.1$. Means, stds and coefficient of variations (CVs) from these simulations are found in Table 5. It is seen that there is a large variation of the means and stds in different replicates. The means varied between 19 and 205 with this small number of replicates and it is evident that much longer simulation periods than 10^6 are needed to get reliable estimates of means and other characteristics of interest.

Table 5 Summary statistics for times-between-FAs in 10 replicates

Replicate:	1	2	3	4	5	6	7	8	9	10	Overall
Mean	44	41	35	103	43	41	87	19	205	32	43.3
Std	56	55	48	122	5	55	109	27	238	47	66.4
CV	1.7	1.8	1.8	1.4	1.7	1.8	1.5	2.0	1.4	2.1	1.91

The large instability is a result of a clustering phenomenon. This can be seen if the series of FAs is approximately treated as a stochastic point process in continuous time. Let N_t denote the number of FAs during time t, measured from an arbitrary time. Then the limiting index of dispersion $I_t = V(N_t)/E(N_t)$, as $t \to \infty$, is $I = 2.1$, indicating a high degree of clustering (cf. Ch. 4.4-4.6 in (6)). In clustered processes short intervals occur more often than longer intervals and the autocorrelation between intervals of lag j is positive and decreases with j. In the present data 20-40 % of all intervals-between-FAs occurred within 1 unit of time and about 2/3 of the intervals were shorter than the mean.

When P_{FA} was set to 0.5 % the clustering tendency was even more pronounced. The means in 10 replicates of length 10^6 now varied between 257 and 34855 and the limiting dispersion index was $I = 2.5$.

As a comparison, consider the distribution of T_{FA} = Times to first FA. This was studied in 10 replicates with the same α_0 as above consisting of 1000 simulated series, each of length 3000 time units. The latter length was enough to secure that at least one FA was obtained in each series. The result is summarized in Table 6. Comparing this table with Table 5 it is apparent that times to first FA is more suitable to use in simulation studies than times-between-FAs.

Table 6 Summary statistics for times to first FA in 10 replicates

Replicate:	1	2	3	4	5	6	7	8	9	10	Overall
Mean	92	92	94	99	91	94	102	92	100	98	95.4
Std	163	178	202	192	190	146	216	151	206	195	185.3
CV	3.1	3.7	4.6	3.8	4.4	2.4	4.5	2.7	4.2	4.0	3.77

3.2 Times to first false alarm and connection with the Weibull distribution

When studying the distribution of times to first FA by simulations it was found that series of length 3000 was enough to secure that at least one FA was observed, provided that P_{FA} was 5 %. However, when the latter probability was smaller it frequently occurred that no FA was observed within the time range of 3000 time units. Due to limited computer resources it was not possible to extend the time range, so a problem with censored data arouse. Luckily it was found that the distribution of times to first FA was well approximated by the Weibull distribution and this facilitated the problem of handling censored data.

Define the random variable T_{FA} = 'Time to first FA', where $T_{FA} \geq 1$ is measured from start of the surveillance and after the sampling and calibration periods. Although T_{FA} is a discrete variable the distribution of $T_{FA} - 1$ agrees well with the continuous two parameter Weibull distribution. According to the latter model the survival function and density of T_{FA} is

$$S(t) = P(T_{FA} > t) = \exp\left\{-\left(\frac{t-1}{\lambda}\right)^{\theta}\right\} \text{ and } f(t) = \frac{\theta}{\lambda}\left(\frac{t-1}{\lambda}\right)^{\theta-1} S(t), \ t \geq 1, \qquad (3a)$$

respectively, and with

$$\text{expectation } \lambda\Gamma(1+1/\theta)+1 \text{ and population median } \lambda(\ln(2))^{1/\theta}+1. \qquad (3b)$$

A simple way to estimate the two parameters θ and λ is to use the fact that

$$\ln(-\ln S(t)) = \theta \cdot \ln(t-1) - \theta \ln(\lambda) \qquad (4)$$

So, by regressing the empirical logarithmic negative log-survival function on logarithmic time one gets the parameter estimates, e.g. by using ordinary least squares (OLS) technique. The agreement between the empirical and the Weibull distributions can be measured by calculating the coefficient of determination for the linear relation in Eq. (4).

To illustrate these procedures, consider 1000 simulated series with $\alpha_0 = 10$, each of length 3000. In each series a sampling period of 10 was used to estimate α_0 followed by a calibration period of 20 after which the surveillance started and the time to first FA was noticed in each series. An alarm was signalled if $S_t > h$, where h was chosen from Table 3 such that P_{FA} was at most 5 % and $k = 1.1$. Linear interpolation was used to determine the final value of h from the estimated α_0 (cf. Section 2.3).

A standardized histogram of times to first FA is shown in Figure 3 together with a fitted Weibull density. The standardized histogram was obtained from the expression $1000 \cdot (\textit{Number of observations in } (t, t+5))/1000 \cdot 5$ and the fitted Weibull density was $1000 \cdot \hat{f}(t)$, where the latter density is obtained from Eq. (3a) with estimates inserted for the parameters. The OLS estimates of the parameters in Eq. (4) were $\hat{\theta} = 0.645$ and $\hat{\theta} \cdot \ln(\hat{\lambda}) = 2.7263$, i.e. $\hat{\lambda} = 68.6$. Coefficient of determination was $R^2 = 98.6\%$. If the values of $\hat{\theta}$ and $\hat{\lambda}$ are inserted into the expressions for the expectation (E) and median ($q_{.50}$) one gets $\hat{E} = 95.6$ and $\hat{q}_{.50} = 39.9$, close to the observed mean 95.1 and observed median 39.0.

The Weibull distribution was only used as an approximation in order to simplify the analysis when data are censored. (Censoring occurred when P_{FA} was below 5 %.)

Figure 3 Frequency distribution of times to first FA together with fitted Weibull density.
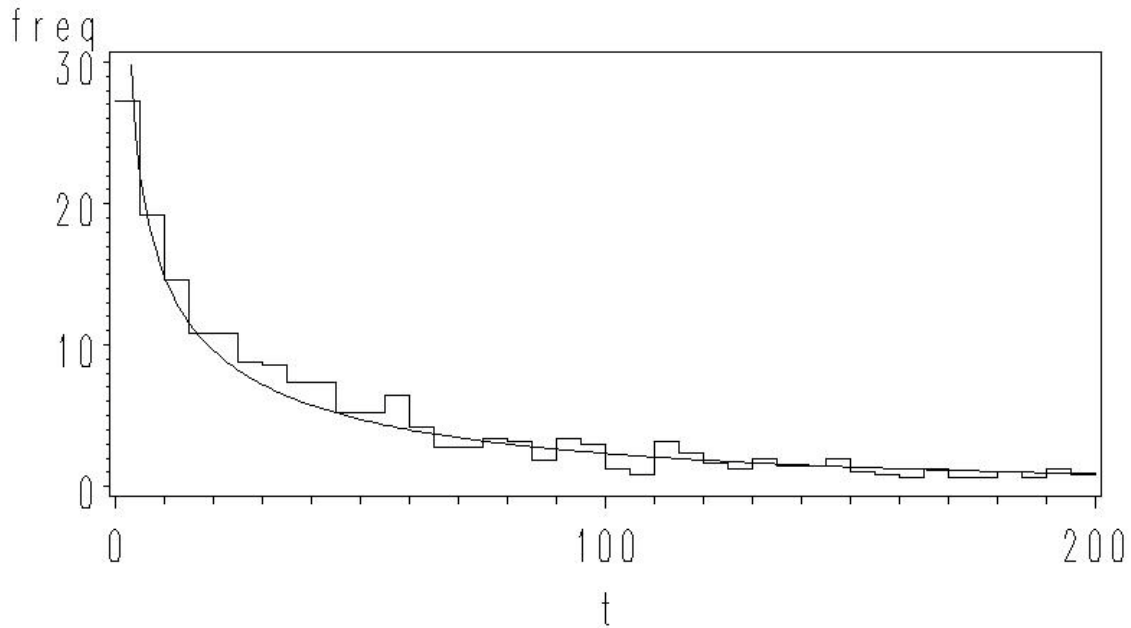
The assumption of a Weibull distribution was merely used to overcome the problem with censored observations. Although the agreement with the latter distribution turned out to be fairly good, it is far from perfect. The random variable T_{FA} should in fact have the density (cf. p. 62 in (6))

$$f(t) = S_X(t)/E(X) \to 1/E(X), \text{ as } t \to 0$$

where the random variable X is the times-between-FAs. The latter behaviour is not in agreement with the Weibull density since, for $\theta < 1$ the density tends to infinity and not to a constant as t tends to zero. Despite this lack of agreement noticed for small times the Weibull model was used as an approximation.

3.3 Times to first motivated alarm

In order to study the distribution of times to first motivated alarm (T_{MA}), 10 000 simulated outbreaks were generated after the sampling and calibration periods (cf. Sect. 2.2). After each outbreak an observed value of T_{MA} was obtained. The constant mean of the Poisson distribution was changed at time τ from α_0 to

$$\alpha_1 = \alpha_0 \exp\{\beta(i - \tau + 1)\},\ i \geq \tau \qquad (5)$$

with $\beta = 0.1, 0.2, 0.3$ and where $\alpha_0 = 10$ as usual. The distribution of T_{MA} is illustrated in Figure 4 where it is evident that the latter distribution is heavily dependent on the parameter β. The distribution of T_{MA} is much more concentrated than that of T_{FA}, suggesting that means and medians can be estimated from relatively short simulated series without having to worry about censored observations.

The ability to detect an outbreak is dependent on the relative increase of the mean, $RI = \alpha_1 / \alpha_0 - 1$. Table 7 shows the values of RI that are obtained 1-5 time units after the outbreak for various β in Eq. (5). Here RI varies between 0.11 and 3.48. This interval covers well the RIs that are obtained for the data that are studied in Section 5 of this paper (about 0.45 to 0.7). The RIs also cover those used by Perry and Pignatiello[18] (0.25-0.75) and by Han et al.[9] (1.25-2.0). In the latter two cases only jump changes were considered. Example of outbreaks with much larger RIs can be found in (9). Here continuous changes were considered, and for 1-5 time units after the outbreak RI increased from 2 to 80 for a data set denoted LDI (laboratory diagnosed influenza cases) and from 28 to 432 for the data set ILI (patients with influenza-like symptoms).

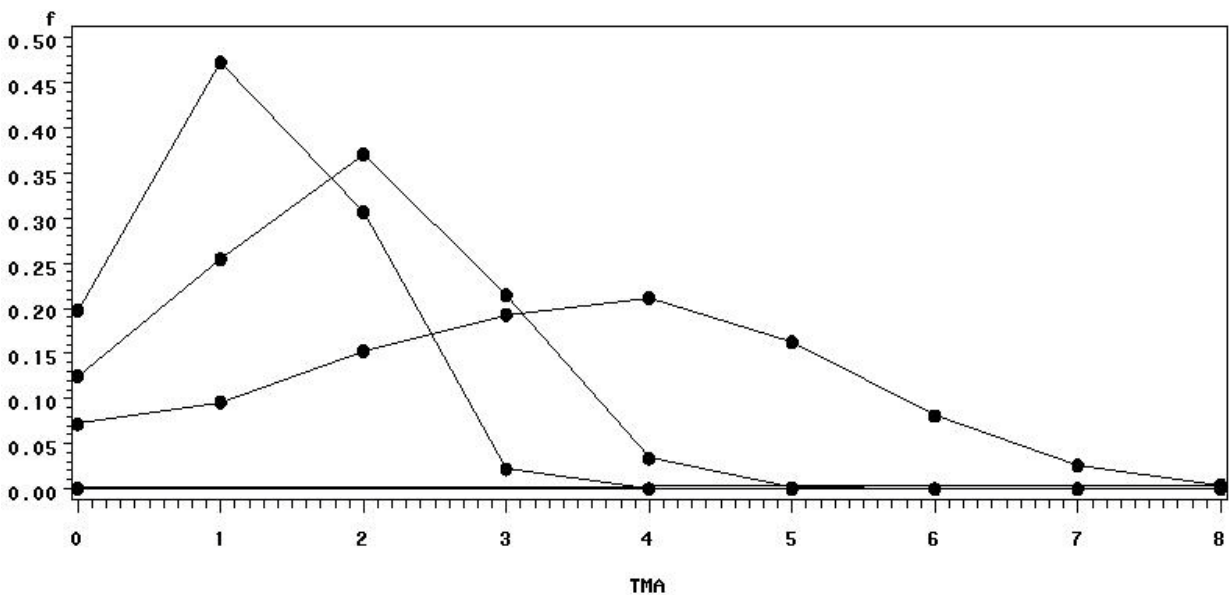**Figure 4** Relative frequencies of times to first motivated alarm (T_{MA}) when β in (5) takes the values 0.1, 0.2 and 0.3.

Table 7 Relative increase of the mean in Eq (5), $RI = \alpha_1 / \alpha_0 - 1$, for various β and times after outbreak.

β :	Time after outbreak				
	0	1	2	3	4
0.1	0.11	0.22	0.35	0.49	0.65
0.2	0.22	0.49	0.82	1.23	1.73
0.3	0.35	0.82	1.46	2.32	3.48

3.4 Sensitivity, specificity and prediktive values

Specificity $(Spec) = 1 - P_{FA}$ was fixed in advance to 95 %, 99 % and 99.5 % and sensitivity (*Sens*) was calculated as

$$Sens = P(S_t > h | \text{outbreak at } t = \tau) \text{ for } t = \tau...\tau + d$$

Here the alarm limit *h* is determined for given values of 1-*Spec* in Table 4, and in the tables in Appendix B for the statistic suggested by Rossi *et al*. The integer *d* was at most 4.

The measures *PPV* and *NPV* are more complex since they depend on the probability of having an outbreak, say P_{OUT}. One way to deal with the problem is to express the two measures in terms of *Sens* and *Spec*. From Baye's theorem one easily obtains the relations

$$PPV = \left[1 + \frac{(1 - Spec)}{Sens} \cdot \frac{(1 - P_{OUT})}{P_{OUT}} \right]^{-1}, \quad NPV = \left[1 + \frac{(1 - Sens)}{Spec} \cdot \frac{P_{OUT}}{(1 - P_{OUT})} \right]^{-1} \quad (6)$$

For given values of *Sens* and *Spec*, *PPV* is a monotonously increasing function of P_{OUT} while *NPV* is monotonously decreasing. It is also seen that if two methods have the same values of *Sens* and *Spec*, then they must have the same predictive values. In practice it may be useful to base the calculation of the predictive measures on earlier estimates of P_{OUT}.

4. Results

4.1 Parameter settings in the Monte Carlo simulations and estimation methods

Time to first false alarm (T_{FA}) before outbreak and time to first motivated alarm (T_{MA}) after outbreak were studied for the reference values $k = 1.1, 1.3, 1.5$ and when P_{FA} was 0.5 %, 1 % and 5 %, thus nine combinations in total. As described earlier, histograms of T_{FA} suggested a strictly decreasing distribution with large variance that was closely connected with the Weibull density. Histograms of T_{MA} showed unimodal distributions with small variance.

For T_{FA} the mean and median was estimated by using four methods. One method based on moments, called MOM, simply calculated the sample mean and sample median and from the latter estimates of the Weibull parameters were obtained from Eq. (3b). Another method, called OLS, used the relation in Eq. (4) to estimate the Weibull parameters and from these the expectation and median were obtained by inserting the parameter estimates into Eq. (3b). MOM and OLS could only be used when P_{FA} was 5 %. For lower values of P_{FA} a certain proportion of T_{FA} was never observed since no FA occurred during the simulated range of 3000 time units. In this case the Weibull parameters were estimated by using the Maximum Likelihood method for censored data[5], here called ML, and also a method based on quantiles[13] which simply will be called Q. The relative efficiency of the latter Q-method is at least 50 % compared with ML, but it has the advantage that the method is more reliable (anomalous estimates were never obtained) and can be used with computers having less capacity. (See (13) for details.)

4.2 Times to first false alarm

Table 8 summarizes the estimates from the simulated series. Both median and mean of T_{FA} decreased with increasing values of P_{FA}, and also with increasing reference value k. As expected, the location parameter λ was closely related to median and mean, but it is quite remarkable that the shape parameter θ remained almost constant. Although there is some variation between the estimates obtained with the different methods, it is hard to find any systematic deviations.

Table 8 Estimates of median T_{FA}, mean T_{FA} and Weibull parameters θ and λ, obtained by up to four estimation methods. Results are only presented in cases where the methods gave reliable estimates. An average value has been added in the table to simplify conclusions from the study.

P_{FA}	k	Method	Median	Mean	θ	λ
0.5 %	1.1	Q	3000	7440	0.64	5330
	1.3	ML	1065	2421	0.67	1836
		Q	1146	2913	0.63	2053
		Average	*1106*	*2667*	*0.65*	*1945*
	1.5	ML	635	1320	0.71	1060
		Q	670	1465	0.69	1140
		Average	*653*	*1393*	*0.70*	*1100*
1.0 %	1.1	ML	765	1742	0.67	1320
		Q	755	2085	0.61	1391
		Average	*761*	*1914*	*0.64*	*1355*
	1.3	ML	339	746	0.68	577
		Q	360	774	0.69	609
		Average	*350*	*760*	*0.69*	*593*
	1.5	ML	228	488	0.69	384
		Q	230	498	0.69	390
		Average	*229*	*493*	*0.69*	*387*
5.0 %	1.1	MOM	37.0	92.0	0.64	66.1
		OLS	39.3	87.5	0.67	65.9
		ML	38.0	95.4	0.63	66.4
		Q	37.3	67.4	0.79	59.2
		Average	*37.9*	*85.6*	*0.68*	*64.4*
	1.3	MOM	29.0	66.6	0.67	50.0
		OLS	30.3	59.0	0.73	48.1
		ML	25.9	66.6	0.62	45.1
		Q	28.6	51.0	0.80	45.1
		Average	*28.5*	*60.8*	*0.70*	*47.1*
	1.5	MOM	23.0	49.4	0.69	38.8
		OLS	21.5	54.3	0.62	37.0
		ML	21.5	53.8	0.63	36.9
		Q	23.7	43.5	0.78	37.9
		Average	*22.4*	*50.3*	*0.68*	*37.7*

The results from the simulations in Table 8 can be compressed further by finding a functional relation between the mean of T_{FA}, say E_{FA}, as a dependent variable (the averages in Table 8) and P_{FA} and k as independent variables. The following relation was obtained (notice that P_{FA} is given in %)

$$E_{FA} = a_{FA} \cdot P_{FA}^{\,b_{FA}} \cdot k^{\,c_{FA}},$$
$$\text{with } a_{FA} = 2471, b_{FA} = -1.6683, c_{FA} = -3.8504 \tag{7}$$

The coefficient of determination (R^2) for the linearized relation in Eq. (7) was 98 %. The expression in Eq. (7) will be compared with the corresponding expression for the expectation of T_{MA}.

4.3 Times to first motivated alarm

The mean times to first motivated alarm (E_{MA}) was studied for the nine combinations of k and P_{FA} mentioned before, but also for $\beta = 0.1, 0.2, 0.3$ in Eq. (5), where a large value of β indicates a more distinct outbreak. The result is shown in Table 9. As might be expected E_{MA} decreased with increasing β and P_{FA}, but also with increasing k.

Table 9 Mean times to first motivated alarm after outbreak (E_{MA}) for some values of the exponential change parameter β. P_{FA} and k is the false alarm probability and reference value, respectively.

		Mean		
P_{FA}	k	$\beta = 0.1$	$\beta = 0.2$	$\beta = 0.3$
0.5%	1.1	5.53	3.06	2.11
	1.3	5.17	2.78	1.86
	1.5	5.10	2.69	1.79
1.0%	1.1	4.91	2.68	1.81
	1.3	4.64	2.48	1.65
	1.5	4.56	2.40	1.57
5.0%	1.1	3.35	1.78	1.15
	1.3	3.16	1.69	1.06
	1.5	3.11	1.65	1.04

A power model fitted to the 27 estimated mean times gave the following result

$$E_{MA} = a_{MA} \cdot P_{FA}^{b_{MA}} \cdot k^{c_{MA}} \cdot \beta^{d_{MA}},$$
$$\text{with } a_{MA} = 0.62, b_{MA} = -0.236, c_{MA} = -0.443, d_{MA} = -0.912 \tag{8}$$

R^2 for the linearized relation in (8) was 99 %.

4.4 Conclusions about expected times

The relations in Eq. (7) and Eq. (8) can be used in several ways. In practice, one wants E_{FA} to be large and E_{MA} to be small. However, it is seen that both E_{FA} and E_{MA} decreases when P_{FA} and k increase, so the latter relations are of less value for finding optimal values of P_{FA} and k. Consider the ratio E_{FA}/E_{MA}, which should be large. From Eq. (7) and Eq. (8) it follows that

$$E_{FA}/E_{MA} = 3985 \cdot \beta^{0.912} \cdot P_{FA}^{-1.225} \cdot k^{-3.614} \tag{9}$$

In the last expression the ratio is large when: (i) β is large (being beyond our control), (ii) P_{FA} is small and (iii) k is small. The fact that k should be chosen small makes the choice of a proper value of k a bit complicated since it was demonstrated in Sect.2 that a small value of k has as a consequence that series of S_t needs longer time to reach a stationary state. Some values of E_{FA}/E_{MA} are seen in Table 10. From Eq. (9) or Table 10 it may be difficult in practice to determine proper values of P_{FA} and k because it may hard to find a reasonable value of the ratio. However, Table 10 may yet give some guide lines. E.g. it can be concluded that the ratio is roughly three times larger for $k = 1.1$ than for $k = 1.3$ for given values of P_{FA} and β.

An alternative approach would be to first specify an acceptable value of E_{MA} (e.g. 2 or 3 time units), a value of k from the thumb rule in Section 2.2 and an assumed range of β. Then the effects of the chosen parameter values on E_{FA} and P_{FA} can be studied in the following relations that are obtained from Eq. (7) and Eq. (8)

$$E_{FA} = 72522 \cdot E_{MA}^{7.069} \cdot k^{-0.7188} \cdot \beta^{-6.447}, \quad P_{FA} = 108 \cdot E_{FA}^{-0.60} \cdot k^{-2.308} \tag{10}$$

Instead of the exponential parameter β one may consider the relative change of the mean process t time units after the outbreak that was introduced in Ch. 2.2, $RI(t) = \exp(\beta t) - 1$, or $\beta = t^{-1} \ln(1 + RI(t))$. E.g. a relative increase of 10% one time unit after the outbreak corresponds to a value of β of about 0.1. In Eq. (10) it is clear that a small increase of E_{MA} will give rise to a much larger increase of E_{FA}. E.g. an increase of E_{MA} from 2 to 3 will give a value of E_{FA} that is $(3/2)^{7.069} = 17.6$ times larger.

Table 10 Values of the ratio E_{FA}/E_{MA} for some parameters.

$P_{FA}(\%)$	k	$\beta = 0.1$	$\beta = 0.2$	$\beta = 0.3$
		E_{FA}/E_{MA}		
0.5	1.1	808	1520	2200
	1.3	442	832	1204
	1.5	263	496	718
1.0	1.1	346	651	942
	1.3	189	356	515
	1.5	113	212	307
5.0	1.1	48	91	131
	1.3	26	50	72
	1.5	16	30	43

4.5 Comparison with expected times for the CUSUM statistic suggested by Rossi *et al*

The performance of the procedures based on the two statistics $Z^{(3)}$ and Z in (2) was compared in simulations with $\alpha_0 = 10$ and $\beta = 0.1, 0.2, 0.3.$. The *h*-values for the CUSUM method based on $Z^{(3)}$ in Eq. (2) were determined in two ways. From actual simulated distributions (cf. table in Appendix B) in the same way as the *h*-values for Z in Table 4 were obtained. The CUSUM statistic obtained in this way is denoted $S_t^{(3)}(S)$. The corresponding statistic based on *h*-values calculated under the assumption that $Z^{(3)}$ is a standard normal variate (cf. the table in Appendix B) is denoted $S_t^{(3)}(N)$. The CUSUM statistic based on Z is denoted S_t.

The expected times to first false alarm (E_{FA}) and to first motivated alarm (E_{MA}) are presented in Table 11a and Table 11b, respectively. In Table 11a it is seen that S_t produces the longest expected times to first false alarm. The values for $S_t^{(3)}(S)$ and $S_t^{(3)}(N)$ are much shorter, especially if P_{FA} and k are small. The expected times to first motivated alarm in Table 11b are smallest for $S_t^{(3)}(N)$ is to be expected since the statistic tends to trigger alarms more frequently. The statistic S_t has the longest expected times to first motivated alarm. However, the difference is small and should be put in relation to the huge differences that were noticed for expected times to first false alarms.

Due to the poor values of E_{FA} for the statistic $S_t^{(3)}(N)$ the latter will not be considered in the sequel.

Table 11a Expected times to first false alarm (E_{FA}) for CUSUMs based on the statistic S_t and the two statistics $S_t^{(3)}(S)$ and $S_t^{(3)}(N)$. (See text.) Relative differences from the results obtained for S_t are shown in parentheses.

P_{FA}	k	S_t	$S_t^{(3)}(S)$	$S_t^{(3)}(N)$
0.5%	1.1	7440	695 (-91%)	405 (-95%)
	1.3	2667	681 (-74%)	334 (-87%)
	1.5	1393	570 (-59%)	281 (-80%)
1.0%	1.1	1914	499 (-71%)	235 (-88%)
	1.3	760	445 (-41%)	200 (-74%)
	1.5	493	353 (-28%)	167 (-66%)
5.0%	1.1	85.6	71.8 (-16%)	45.2 (-47%)
	1.3	60.8	55.0 (-10%)	36.7 (-40%)
	1.5	50.3	45.7 (-9%)	35.7 (-29%)

Table 11b Expected times to first motivated alarm (E_{MA}) for CUSUMs based on the same statistics as in Table 11a.

P_{FA}	k	β:	S_t 0.1	0.2	0.3	$S_t^{(3)}(S)$ 0.1	0.2	0.3	$S_t^{(3)}(N)$ 0.1	0.2	0.3
0.5%	1.1		5.5	3.0	2.1	5.0	2.7	1.9	4.2	2.3	1.5
	1.3		5.1	2.8	1.8	4.9	2.7	1.8	4.3	2.3	1.5
	1.5		5.1	2.7	1.8	4.9	2.6	1.7	4.3	2.3	1.5
1.0%	1.1		4.9	2.7	1.8	4.6	2.5	1.7	3.9	2.1	1.3
	1.3		4.6	2.5	1.6	4.5	2.4	1.6	3.9	2.1	1.3
	1.5		4.6	2.4	1.6	4.5	2.4	1.5	4.0	2.1	1.3
5.0%	1.1		3.2	1.7	1.1	3.2	1.7	1.1	2.8	1.5	1.0
	1.3		3.1	1.7	1.0	3.1	1.7	1.0	2.8	1.5	1.0
	1.5		3.1	1.6	1.0	3.1	1.6	1.0	2.8	1.5	0.9

4.6 Sensitivity and predictive values

The sensitivity (*Sens*) of CUSUM based on S_t and $S_t^{(3)}(S)$ is shown in Table C in Appendix C. Notice that *Sens* in Section 3.4 was defined for a range of values of d, such that $d = 0$ gives the probability for an alarm at the time of the outbreak and $d = 4$ gives the probability for an alarm within 4 time units after the outbreak.

In Table C it is striking that the two CUSUMs give roughly the same sensitivities. As expected, *Sens* increases with β and d and approaches 1.00 when $\beta = 0.3$ corresponding to a relative increase of the mean that is larger than 1.7 (cf. Table 7). It is also seen that *Sens* increases slightly with P_{FA} but is rather unaffected by k.

From Table C the predictive values *PPV* and *NPV* can be calculated by using the relations in Eq. (7). To take some examples, consider the case when $P_{FA} = 0.5\%$ (i.e. *Spec* = 0.995), $k = 1.1$ and $\beta = 0.3$. Using $d = 0$ will give *Sens* = 0.03 in Table C. Then, as the probability of an outbreak increases from 0.1 to 0.9, *PPV* increases from 0.40 to 0.98 and *NPV* decreases from 0.90 to 0.10. By instead using $d = 2$ one obtains *Sens* = 0.85. As the probability of an outbreak increases from 0.1 to 0.9, *PPV* increases from 0.95 to 1.00 and *NPV* decreases from 0.98 to 0.42.

5. Two examples

5.1 Outbreak of sick-listening among women in a Swedish municipality

During the years 1995-2002 there was an unexpected large increase of the number of long-termed (more than 60 days) sick-listed persons in Sweden. This was noticed in all 290 municipalities of the country, but with various times of onsets and with various intensities. The increase was a result of an increased number of new cases (incidence rate) and in some regions also a result of prolonged durations of the sickness period. Further details about these issues can be found in a paper by Nilsson and Jonsson[19]. In order to exemplify the procedures described in earlier sectionsr, data for women in the municipality of Sollefteå (roughly 20 000 inhabitants, both sexes) will be used.

Figure 5 shows the development of the number of new cases of long-termed sick-listened women per month from January 2 1996 to December 31 2001. It is seen that the development is quite stable up to about $t = 45$ (September 1999), but after this the number of cases per month increases from about 20 to more than 60.
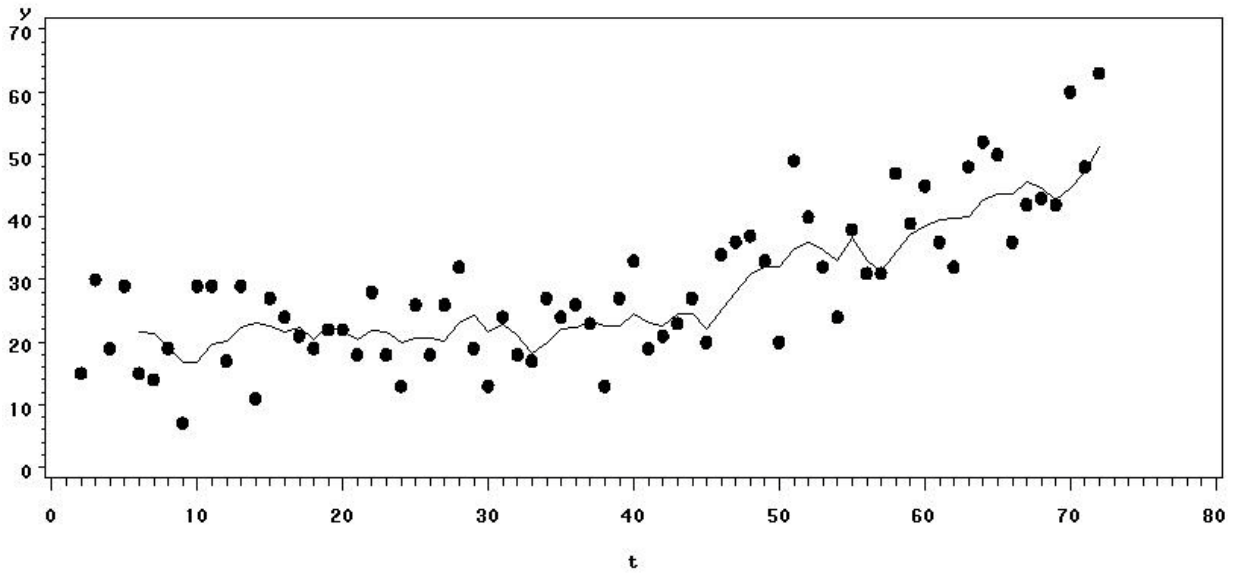
Figure 5 Dots representing the number of new cases of long-termed sick-listed women in Sollefteå, per month January 2 1996 (*t* = 2) to December 31 2001 (*t* = 72). The number of cases is plotted together with a moving-average curve.

The average number of cases during the first 10 months was 20.6. By using the latter value, the alarming boundary *h* for the statistic S_t can be found by linear interpolation in Table 4. The corresponding alarming boundaries for the statistics $S_t^{(3)}(S)$ and $S_t^{(3)}(N)$ (cf. Ch. 4.5), *h*(*S*) and *h*(*N*) respectively, are obtained similarly from the table in Appendix B. In the latter table *h*(*S*) is determined from simulations and *h*(*N*) is determined from the assumption that $Z^{(3)}$ is a standard normal variate. The three types of alarming boundaries are compared in Table 12. It is seen that the *h*-values are smallest for the two approaches due to Rossi *et al*. On the other hand, the CUSUM statistics are somewhat lower for the Rossi approach, as can be seen in Figure 6 below.

P_{FA} (%)	k	h	h (S)	h (N)
5	1.1	1.18	0.98	0.74
	1.3	0.78	0.64	0.45
	1.5	0.47	0.35	0.21
1	1.1	2.74	2.08	1.52
	1.3	1.91	1.60	1.19
	1.5	1.46	1.19	0.92
0.5	1.1	3.56	2.59	1.85
	1.3	2.48	2.02	1.49
	1.5	1.87	1.53	1.19

Table 12 Three different alarming boundaries (h) for some values of the probability of a false alarm at a single time point (P_{FA}(%)) and k in Eq. (1).

To demonstrate the performance of the three approaches, consider the case were it is required that $P_{FA} = 1\%$ and one wants to detect a relative increase of about 50 %, i.e. $RI =$ 0.50. Since $\hat{\alpha}_0 = 20.6$ one gets from Table 3 $k \approx 2.27 \cdot 0.50 \approx 1.1$ and from this the value of h is determined from Table 12. The result is seen in Figure 6. Here the CUSUM statistics based on Z in (2) are equal or somewhat larger than those based on $Z^{(3)}$. Also the alarming boundaries are larger, $h = 2.74$, compared with $h(S) = 2.08$ and $h(N) = 1.52$. Already at t = 13 CUSUM(N) gives an alarm, while CUSUM(S) is very close to 2.08 but does not alarm. At t =22 both of the latter CUSUMs give alarm and at $t = 27$, 28 and 40 all three CUSUMs alarm. The next times that alarms are given from all three CUSUMS are from $t = 47$ and onwards.

$RI = 0.50$ corresponds to a value of the exponential parameter β that is 0.40 1 time unit after outbreak and 0.20 2 time units after outbreak (Cf. Ch. 4.4.). From Eq. (8) the expected time to first motivated alarm (E_{MA}) is 1.4 for $\beta = 0.4$ and 2.6 for $\beta = 0.2$.

It is interesting to compare the alarms given by the three CUSUMs with the actual development plotted in Figure 6. There is a great variability in the number of cases from month to month, so just a few high values may not be an indication of a trend. In Figure 6 it is obvious that clear signs of an outbreak can be seen from about $t = 46$ and this would not have been possible to detect from a moving average curve at that time. The fact that all three CUSUMs alarmed after this time point (with some exceptions) is to be expected, but perhaps more interesting is to study what happened before the outbreak. CUSUM(N) was very keen on alarming before the outbreak (6 times), and to less extent also CUSUM(S). These results are

in agreement with the simulation studies in Ch. 4. Moore difficult to explain is that the CUSUM presented in this paper alarmed three times before the outbreak. Here only the case $P_{FA} = 1\%$ was considered. With $P_{FA}(\%) = 0.5$ alarm was given only once before the outbreak .
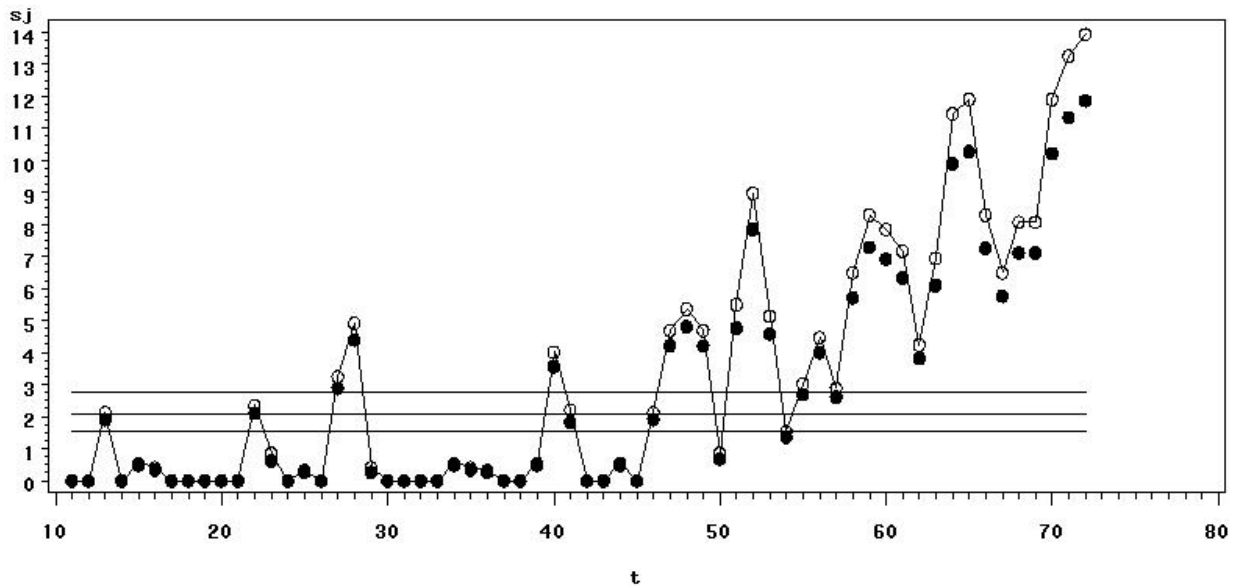

Figure 6 CUSUM statistics plotted against time. Circles represent the statistics suggested in the present paper while dots are the statistics suggested by Rossi et. al. The three horizontal lines are (from upper to lower) the alarming bounds given by h, $h(S)$ and $h(N)$.

5.2 Detection of trends and seasonal variability in outbreaks of Chlamydial infection

Chlamydia infection is a common sexually transmitted infection in humans. In Sweden Chlamydia is classified as a disease that is dangerous to the public and all cases are obliged to be reported to the authority (Swedish Institute for Disease Control, SMI). After the millennium an increase of the incidence in Chlamydia was noticed in all Swedish counties[27]. The increase was most pronounced in counties where tourists are crowded during the summer vacations. A typical example is the county of Halland at the Swedish west coast, where several popular beaches and dancing places are situated.

The number of reported cases of Chlamydia per month in Halland during the period January 1 2000 ($t = 1$) to December 31 2007 ($t = 96$) is shown in Figure 7. The development seems to be quite stable up to about $t = 44$ (August 2004), even if smaller clusters of outbreaks can be distinguished earlier. After this there is an increase from roughly 50 cases per month to about 130. The moving average curve in Figure 7 reveals a weak seasonal variability with 8 peaks occurring during the 8-year period, although earlier peaks are less pronounced.
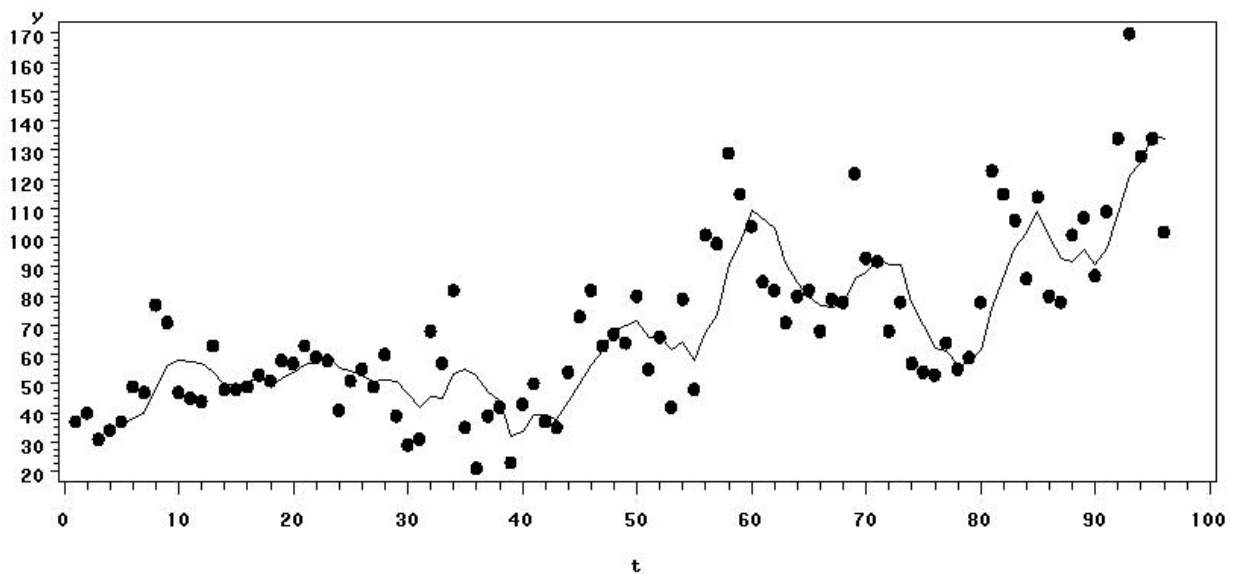

Figure 7 Dots represent the number of new cases of Chlamydia in Halland per month January 1 2000 ($t = 1$) to December 31 2007 ($t = 96$). Cases are plotted together with a moving average curve.

Average value during the first 10 months was 47.0. From this estimate the alarming boundaries can be obtained by linear interpolation in Table 3 and the tables in Appendix B. With $P_{FA} = 0.5\%$ and $RI = 0.4$ one gets $k \approx 1.3$ and the boundaries $h = 2.30$, $h(S) = 1.96$ and $h(N) = 1.49$. The CUSUM statistics are plotted in Figure 8. The first alarm appears at $t = 32$ for CUSUM(N) and at $t = 34$ for CUSUM and CUSUM(S). Further alarms are given repeatedly, especially close to the peaks of the periodic outbreaks.
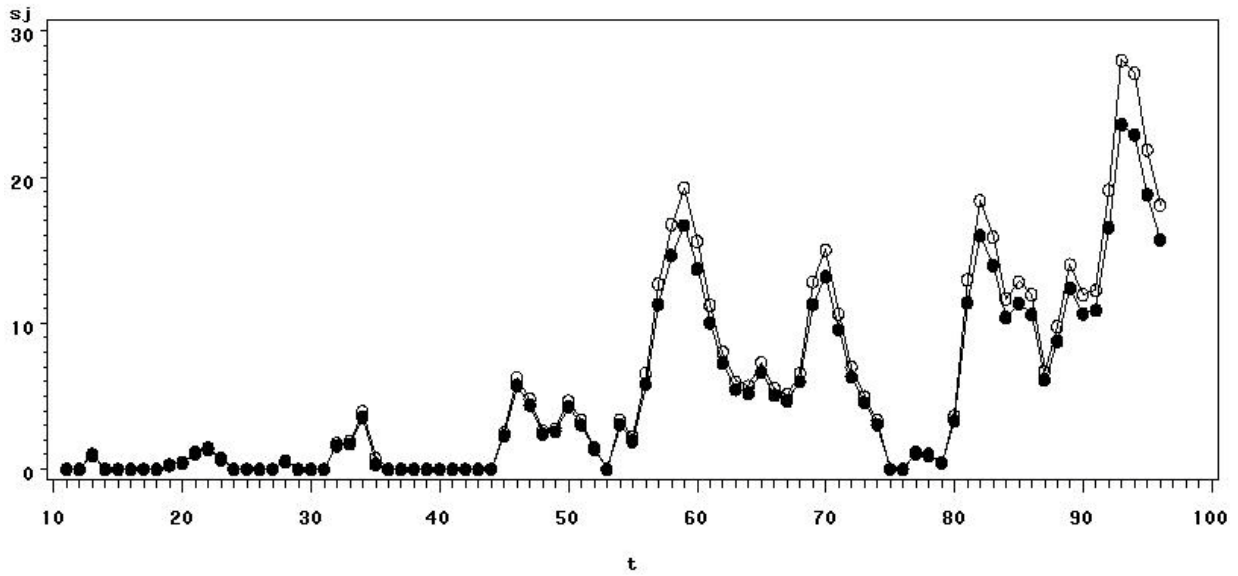
Figure 8 CUSUM statistics plotted against time where circles represent the statistic suggested in the present paper and dots are the statistics suggested by Rossi *et al.* Alarming boundaries are not shown in the figure.

It may be of some interest to study the peaks in Figure 8. Table 13 gives the locally (within one year) largest peak heights per year and the times at which the peaks were reached. If one compares Figure 7 with Figure 8 one can notice that the peaks of the CUSUM statistics appear constantly earlier than those of the 5-point moving average curve. E.g. the latter curve has a local peak at $t = 85$, whereas the CUSUM statistic has a peak three months earlier at $t = 82$. This illustrates again that although moving average curves may be a rough way of finding trends in data, it is less suitable for detecting outbreaks.

Table 13 Largest local peak heights for the CUSUM statistic presented in this paper during the years 2001-2007, and time at which the peaks were reached.

Year	2001	2002	2003	2004	2005	2006	2007
Peak height	1.47	3.95	6.78	19.3	15.0	18.4	28.0
t (month)	22 (Oct)	34 (Oct)	46 (Oct)	59 (Nov)	70 (Oct)	82 (Oct)	93 (Sept)

The discovery of a seasonal variation of Chlamydia cases seems not to have been reported earlier. In a paper by Rolfhamre[23] Chlamydia is mentioned as an example of a disease without outbreaks, in contrast to diseases with periodic outbreaks. This seems no longer to be true.

6. Concluding discussion

When determining the alarming boundary h it has become a common practice to first focuse on the relation between the expected times to first false and motivated alarms, E_{FA} and E_{MA}, respectively. The latter times have also being denoted average run lengths (ARLs). This seems not to be particularly useful for predicting outbreaks of medical health events for at least two reasons. One is that expected values as a measure in this case give a distorted picture of reality. In fact, the simulations carried out in Section 4 showed that 70 % - 80 % of all first false alarms occurred before the expected value. Similar criticism has been given in other papers (see e.g. Section 4 in (18)). A second reason is that it may be hard in practice to determine a desirable balance between E_{FA} and E_{MA}.

In this paper the focus instead has been primarily on P_{FA}, the probability of a false alarm at an arbitrary time point. The latter was set to 5 %, 1 % and 0.5 %, corresponding to a specificity of 95 %, 99 % and 99.5 %, respectively. Compared with other studies, values of the specificity of 95 % or more might be considered as extremely large. E.g. Choi et al use values of about 74 % - 95 %[4]. The reason for demanding high specificity is to avoid an 'the boy who cried wolf on'- effect, i.e. the fact that repeated false alarms may undermine trust in warnings issued. In this case there is a difference between what levels of specificity to choose when CUSUM is used in public medicine or in other areas such as industrial process control.

A stumbling block in this type of studies is determination of the reference value k and of the alarm boundary h. Here various ways of doing this have been outlined. One approach starts with a given P_{FA} and assessing a relative increase of the mean, RI, that one is aiming to detect. The reference value k is then determined from Table 3 by using an estimate of α_0, the

mean before outbreak. The alarm boundary is finally obtained from Table 4. Alternatively, as mentioned in Section 4.7, one can focus on the relation between expected time to first false alarm and motivated alarm, E_{FA} and E_{MA}, respectively. The latter approach seems to be less useful in practice since it may be hard to set up a proper balance between E_{FA} and E_{MA}. In any case one should be aware of the danger of choosing k too small since in that case the CUSUM procedure may need a longer time to reach stationarity, as was described in Section 2.2.

Throughout the paper the mean level before outbreak, α_0, was estimated by using values from just 10 time points, which may seem a bit paltry. This was done merely for convenience when comparing difference methods. In practice estimates of α_0 can be computed sequentially during much longer periods, provided that you can be sure that the base line level is unchanged.

It has been shown that the CUSUM procedure based on Z in Eq. (2) has about the same expected times to motivated alarm and the same sensitivity as the procedure suggested by Rossi et al. The difference is that the former increases the expected time to first false alarm by up to 90 % (cf. Table 11a) compared with the method of Rossi et al. . So, the procedure based on Z in Eq.(2) that has been presented in this paper should be used if it is important to keep the frequency of false alarms at a low level.

Appendix

A. Approximate expectations of $Z^{(1)}, Z^{(2)}, Z^{(3)}$ and Z

Consider the approximation $E(g(X)) \approx g(\mu) + g''(\mu)/2 \cdot \sigma^2$, where $\mu = E(X)$ and $\sigma^2 = V(X)$, that follows from a Taylor series approximation (cf. p. 328 in (2)). Let α_0 and α_1 be the expectations of Y before and after the outbreak, respectively, and put $R = \alpha_1 / \alpha_0$. The mean and variance of the estimator $\hat{\alpha}_0$ based on n independent observations before the outbreak is α_0 and α_0 / n, respectively.

$$Z^{(1)} = \frac{Y - \hat{\alpha}_0}{\sqrt{\hat{\alpha}_0}} = Y\hat{\alpha}_0^{-\frac{1}{2}} - \hat{\alpha}_0^{\frac{1}{2}}. \text{ Here } E(\hat{\alpha}_0^{-1/2}) \approx \alpha_0^{-1/2}\left(1 + 3\alpha_0^{-1}/8n\right) \text{ and}$$

$$E(\hat{\alpha}_0^{1/2}) \approx \alpha_0^{1/2}\left(1 - \alpha_0^{-1}/8n\right), \text{ so } E(Z^{(1)}) \approx (R-1)\sqrt{\alpha_0} + \frac{3R+1}{8n} \cdot \frac{1}{\sqrt{\alpha_0}}, \text{ which reduces to}$$

$$\frac{1}{2n\sqrt{\alpha_0}} \text{ when } R = 1, \text{ i.e. before the outbreak.}$$

$$Z^{(2)} = 2\left(\sqrt{Y} - \sqrt{\hat{\alpha}_0}\right) \text{ has } E\left(Z^{(2)}\right) \approx 2\left[\left(\alpha_1^{\frac{1}{2}} - \frac{1}{8}\alpha_1^{-\frac{1}{2}}\right) - \left(\alpha_0^{\frac{1}{2}} - \frac{1}{8n}\alpha_0^{-\frac{1}{2}}\right)\right] =$$

$$= 2\left(\sqrt{R} - 1\right)\sqrt{\alpha_0} - \left(\frac{1}{\sqrt{R}} - \frac{1}{n}\right) \cdot \frac{1}{4\sqrt{\alpha_0}}, \text{ which reduces to } -\left(1 - \frac{1}{n}\right) \cdot \frac{1}{4\sqrt{\alpha_0}} \text{ when } R = 1. \text{ In}$$

the last case bias does not vanish as $n \to \infty$.

$$Z^{(3)} = \frac{Z^{(1)} + Z^{(2)}}{2} \text{ has } E\left(Z^{(3)}\right) = \left(E(Z^{(1)}) + E(Z^{(2)})\right)/2.$$

The estimator Z in (2) is simply $Z^{(1)} - \frac{1}{2n\sqrt{\hat{\alpha}_0}}$, i.e. an attempt to reduce the bias. From the expressions above it follows that the expected value of Z when $R = 1$ is roughly $-3\alpha_0^{-3/2}/8n$.

B. Alarm boundaries h for CUSUM based on $Z^{(3)}$

For given P_{FA}, k and α_0, values of h were determined for CUSUM based on the exact distribution of $Z^{(3)}$ in the same way as for CUSUM based on Z described in Section 2.3. These are presented in Table B1. Table B2 shows the corresponding values of h under the assumption that $Z^{(3)}$ has a standard normal distribution.

Table B1 Values of h for CUSUM based on the statistic $Z^{(3)}$ in Eq. (2) such that the probability of a false alarm at a single time point is 5 %, 1 % and 0.5 %. Each figure is determined from 100'000 simulations.

α_0	5%			1%			0.5%		
	$k=1.1$	$k=1.3$	$k=1.5$	$k=1.1$	$k=1.3$	$k=1.5$	$k=1.1$	$k=1.3$	$k=1.5$
5	1.02	0.69	0.40	2.27	1.74	1.37	2.85	2.19	1.75
10	1.01	0.67	0.38	2.17	1.66	1.28	2.69	2.12	1.67
15	1.00	0.66	0.37	2.12	1.63	1.24	2.64	2.06	1.60
20	0.99	0.65	0.36	2.09	1.62	1.20	2.60	2.04	1.54
30	0.98	0.64	0.35	2.08	1.59	1.19	2.59	2.00	1.53
40	0.97	0.64	0.35	2.07	1.58	1.18	2.58	1.98	1.50
50	0.96	0.63	0.34	2.06	1.57	1.18	2.56	1.95	1.49
75	0.95	0.62	0.34	2.05	1.52	1.17	2.54	1.92	1.47
100	0.93	0.60	0.31	1.97	1.47	1.11	2.43	1.85	1.44
150	0.93	0.60	0.30	1.96	1.46	1.10	2.43	1.85	1.43
200	0.93	0.60	0.30	1.96	1.46	1.10	2.43	1.85	1.43

Table B2 Values of h as in Table B1, but where $Z^{(3)}$ is assumed to have a standard normal distribution.

5%			1%			0.5%		
$k=1.1$	$k=1.3$	$k=1.5$	$k=1.1$	$k=1.3$	$k=1.5$	$k=1.1$	$k=1.3$	$k=1.5$
0.736	0.453	0.207	1.523	1.191	0.920	1.853	1.486	1.189

C. Sensitivity of CUSUM based on S_t and $S_t^{(3)}(S)$

Table C

P_{FA}	k	d	β:	S_t			$S_t^{(3)}(S)$		
				0.1	0.2	0.3	0.1	0.2	0.3
0.5%	1.1	0		.01	.02	.03	.01	.01	.03
		1		.02	.08	.26	.02	.08	.27
		2		.07	.37	.85	.07	.37	.85
		3		.17	.81	1.00	.16	.81	1.00
		4		.35	.99	1.00	.35	.99	1.00
	1.3	0		.01	.02	.03	.01	.02	.03
		1		.02	.10	.31	.02	.10	.31
		2		.07	.40	.86	.07	.40	.87
		3		.18	.83	1.00	.18	.82	1.00
		4		.36	.99	1.00	.36	.99	1.00
	1.5	0		.01	.02	.05	.01	.02	.04
		1		.03	.12	.36	.03	.12	.34
		2		.08	.43	.89	.08	.43	.89
		3		.19	.83	1.00	.19	.83	1.00
		4		.38	.99	1.00	.37	.99	1.00
1%	1.1	0		.01	.02	.03	.01	.02	.03
		1		.03	.10	.28	.03	.09	.28
		2		.07	.39	.87	.08	.39	.87
		3		.18	.82	1.00	.18	.82	1.00
		4		.37	.99	1.00	.37	.99	1.00
	1.3	0		.02	.03	.05	.02	.04	.07
		1		.04	.15	.38	.05	.18	.43
		2		.11	.48	.91	.12	.53	.93
		3		.24	.87	1.00	.26	.89	1.00
		4		.44	.99	1.00	.46	.99	1.00

	1.5	0	.02	.04	.08	.02	.04	.08
		1	.06	.19	.47	.06	.18	.46
		2	.14	.53	.93	.14	.53	.93
		3	.28	.89	1.00	.28	.89	1.00
		4	.49	.99	1.00	.48	.99	1.00
5%	1.1	0	.08	.12	.19	.08	.13	.22
		1	.17	.37	.66	.18	.41	.70
		2	.32	.75	.98	.34	.78	.98
		3	.52	.96	1.00	.55	.97	1.00
		4	.73	1.00	1.00	.75	1.00	1.00
	1.3	0	.08	.14	.23	.08	.14	.24
		1	.19	.42	.71	.20	.43	.72
		2	.36	.78	.98	.36	.79	.99
		3	.56	.97	1.00	.57	.97	1.00
		4	.76	1.00	1.00	.77	1.00	1.00
	1.5	0	.09	.15	.24	.09	.15	.24
		1	.21	.43	.73	.21	.45	.73
		2	.37	.79	.99	.38	.80	.99
		3	.57	.97	1.00	.59	.97	1.00
		4	.77	1.00	1.00	.79	1.00	1.00

References

1. Barbujani, G. And Calzolari, E. , Comparison of two statistical techniques for the surveillance of birth defects through a Monte Carlo simulation, *Statistics in Medicine*, **3**, 239-247 (1984).

2. Casella, G. and Berger, R. L., *Statistical Inference*, Duxbury Press (1990).

3. Chen, R. , The relative efficiency of the sets and CUSUM techniques in monitoring the occurrence of a rare event, *Statistics in Medicine*, **6**, 517-525 (1987).

4. Choi, B. Y., Kim, H., Go, U. Y. and Jeong, J. H., Comparison of various statistical methods for detecting disease outbreaks, *Comput. Stat*., DOI 10.1007/s00180-010-0191-7 (2010).

5. Cohen, A.C., Maximum likelihood estimation in the Weibull distribution based on complete and censored samples, *Technometrics*, **7**, 579-588 (1965).

6. Cox, D. R. and Lewis, P.A. W., *The Statistical Analysis of Series of Events*, Methuen, London (1968).

7. Cramer, H., *Mathematical Methods of Statistics* (7:th ed.), Princeton University Press, Princeton (1957).

8. Ewan, W. D. and Kemp, K. W., Sampling inspection of continuous processes with no autocorrelation between successive results, *Biometrika*, **47**, 363-380 (1960).

9. Frisen, M., Andersson, E. and Schiöler, L., Robust outbreak surveillance of epidemics in Sweden, *Statistics in Medicine*, **28**, 476-493 (2009).

10. Han, S.W., Tsui, K-L., Ariyajunya, B. and Kim, S.B. , A comparison of CUSUM, EWMA, and Temporal Scan Statistics for detection of increases in Poisson rates, *Quality and Reliability Engeneering International*, **26**, 279-289 (2010).

11. Hutwagner, L.C., Maloney, E.K., Bean, N.H., Slutsker, L. and Martin, S.M. , Using laboratory-based surveillance data for prevention: An algorithm for detecting Salmonella outbreaks, *Emerging Infectious Diseases*, **3**, 395-400 (1997).

12. Jonsson, R. , A longitudinal approach for constructing β-expectation tolerance limits, *Journal of Biopharmaceutical Statistics*, **13,** 307-324 (2003).

13. Jonsson, R. , Relative efficiency of a quantile method for estimating parameters in censored two-parameter Weibull distributions, *Research Report*, **3**, *2010:* , Statistical Research Unit, Department of Economics, Sweden (2010).

14. Liddell, F. D. K., Simple exact analysis of the standardized mortality ratio, *Journal of Epidemiology and Community Health*, **38**, 85-88 (1984).

15. Lorden, G. Procedures for reacting on a change in distribution, *Annals of Mathematical Statistics*, **42**, 1897-1908 (1971).

16. Lucas, J. M., Counted data CUSUM's, *Technometrics*, **27**, 129-144 (1988).

17. Moustakides, G.V. , Optimal stopping times for detecting changes in distribution, *Annals of Statistics*, **14**, 1379-1387 (1986).

18. Munroue, A.G. , A control chart based on cumulative scores, *Applied Statistics*, **29**, 252-258 (1980).

19. Nilsson, G. and Jonsson, R. , Report from 'Projekt regionala analyser av försäkringsutnyttjandet', *Försäkringskassan*, ISBN-91-631-3739-9, (in Swedish) (2003)

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

21. Perry, M.B. and Pignatiello, J.J. , Estimating the time of step change with Poisson CUSUM and EWMA control charts, *International Journal of Production Research*, **?**, 1-15 (2010).

22. Roberts, S.W. , A comparison of some control chart procedures, *Technometrics*, **8**, 411-430 (1966).

23. Rolfhamre, P. , Outbreak detection of communicable diseases- Design, analysis and evaluation of three models for statistically detecting outbreaks in epidemiological data of communicable diseases, *Master's Thesis in Computer Science*, Department of Numerical Analysis and Computer Science, Royal Institute of Technology, Stockholm, Sweden (2003).

24. Rossi, G., Lampugnani, L. and Marchi, M., An approximative cusum procedure for surveillance of health events, *Statistics in Medicine*, **18**, 2111-2122 (1999).

25. Shewhart, W. A. , Economic Control of Quality of Manifactured Product, *Reinhold Company*, Princeton, N. J. (1931).

26. Sommerville, P. N., Tables for obtaining non-parametric tolerance limits, *Annals of Mathematical Statistics*, **29**, 599-601 (1958).

27. Swedish Institute for Infectious Disease Control, *Statistics for Chlamydial Infection* (in Swedish) (2009).

28. Wilks, S. S., Determination of sample sizes for setting tolerance limits, *Annals of Mathematical Statistics*, **12**, 91-96 (1941).

Research Report

2007:13	Frisén, M., Andersson, E. & Pettersson, K.	Semiparametric estimation of outbreak regression.
2007:14	Pettersson, K.	Unimodal regression in the two-parameter exponential family with constant or known dispersion parameter.
2007:15	Pettersson, K.	On curve estimation under order restrictions.
2008:1	Frisén, M.	Introduction to financial surveillance.
2008:2	Jonsson, R.	When does Heckman's two-step procedure for censored data work and when does it not?
2008:3	Andersson, E.	Hotelling´s T2 Method in Multivariate On-Line Surveillance. On the Delay of an Alarm.
2008:4	Schiöler, L. & Frisén, M.	On statistical surveillance of the performance of fund managers.
2008:5	Schiöler, L.	Explorative analysis of spatial patterns of influenza incidences in Sweden 1999−2008.
2008:6	Schiöler, L.	Aspects of Surveillance of Outbreaks.
2008:7	Andersson, E & Frisén, M.	Statistiska varningssystem för hälsorisker
2009:1	Frisén, M., Andersson, E. & Schiöler, L.	Evaluation of Multivariate Surveillance
2009:2	Frisén, M., Andersson, E. & Schiöler, L.	Sufficient Reduction in Multivariate Surveillance
2010:1	Schiöler, L	Modelling the spatial patterns of influenza incidence in Sweden
2010:2	Schiöler, L. & Frisén, M.	Multivariate outbreak detection
2010:3	Jonsson, R.	Relative Efficiency of a Quantile Method for Estimating Parameters in Censored Two-Parameter Weibull Distributions