GÖTEBORG UNIVERSITY

Department of Statistics

RESEARCH REPORT 1995:1
ISSN 0349-8034

SURVEILLANCE OF RARE EVENTS. ON EVALUATION OF THE SETS METHOD

by

Hrafnhildur Arnkelsdóttir

Statistiska institutionen
Göteborgs Universitet
Viktoriagatan 13
S-411 25 Göteborg
Sweden

SURVEILLANCE OF RARE EVENTS.
ON EVALUATION OF THE SETS METHOD.

By Hrafnhildur Arnkelsdóttir

SUMMARY

Continual surveillance aiming to detect an increased frequency of some rare event is of interest in several different situations in quality control, medicine, economics and other fields. Examples are continual surveillance of defect articles in a production process or surveillance of a business cycle. Surveillance of rare health events in general and especially surveillance of congenital malformations has been a field of unabating interest during the last decades. Since the Thalidomide episode in the early 60's, several registries of congenital malformations are in operation all over the world. The basic idea is that if a 'catastrophe' occurs an alarm should be signalled as soon as possible after the occurrence. A method developed for this situation is the Sets method that focuses on the intervals between events under surveillance, e.g. intervals between successive births of malformed babies. If a previously defined number of such intervals are 'short' an alarm is triggered. The traditional evaluation measure used when discussing the Sets method is the ARL (Average Run Length). Here, evaluation measures such as the probability of a false alarm, the probability of a successful detection and the predictive value of an alarm are derived and discussed for the Sets method. The information provided by these measures is important for the implementation and use of a system of surveillance in practice.

KEY WORDS: surveillance, rare health events, congenital malformations, the Sets method, probability of a false alarm, probability of a successful detection, predictive value.

TABLE OF CONTENTS

The principles of surveillance are as follows: each month new reports are examined continually and clusters of unusual malformations or combinations of malformations noted. Each year a final report is made, this is mainly used for studies of long-term changes and equivalent material is available for the time period since 1965. Quarterly, information is summarised and reported to an international organisation known as the International Clearinghouse for Birth Defect Monitoring Systems (ICBDMS).

1.2. Surveillance in General

Surveillance can be viewed as continual observation in time where the goal is to detect a change in the underlying process as soon as possible after it has occurred. In terms of statistical inference we have a situation with three characteristics: the number of observations is increasing, decisions must be made successively and the "catastrophe" under surveillance could occur at any time. The hypotheses undergo successive changes since at decision point s we are interested in the hypothesis that no change has occurred before s, while at decision point $s+1$ a change before $s+1$ is of interest.

This situation arises in different areas of medicine, for example surveillance of foetal heart rate during labour, regular health controls, post-marketing surveillance of adverse drug reactions and the scope of this paper, i.e. surveillance of congenital malformations. The timeliness and the accurateness of a decision are extremely important in view of the possible consequences of an increase going undetected for a long time, or, alternatively, having frequent false alarms. Each surveillance situation is unique and needs to be evaluated in the light of these factors.

When conducting statistical analyses in this area the fact that repeated decisions will be made and that no fixed hypothesis is of special interest is often ignored. However, several methods that take into consideration the sequential structure of the surveillance situation are available. The literature on surveillance of congenital malformations contains descriptions of a number of suitable methods. The CUSUM method is a well-known method originally developed in the field of quality control, cf. Barbujani (1987), Chen (1987), Gallus et.al. (1986) and Lie et.al. (1991). A method discussed almost exclusively in this context is the Sets method, discussed by e.g. Barbujani (1987), Chen (1987), Gallus et.al. (1986), Lie et.al. (1991) and Sitter (1990). Gallus (1993) discusses the CUSETS method, a method partly based on the same idea as the Sets method. Other methods described in the context of surveillance of congenital malformations are for example Healy´s weighted regression and sequential analysis by Barbujani (1987) among others, SM-scheme by Shore and Quade (1989) and the CUSCORE by Lie et.al. (1991), Radaelli (1992) and Wolter (1987). Radaelli and Gallus (1989) discuss a stopping rule for surveillance of rare health events. Sitter et.al. (1990) present a similar method to the Sets method developed for the surveillance of cancer rates.

Evaluation of the performance of surveillance methods has been discussed in the literature, cf. Frisén (1989) and Åkermo (1993). Several evaluation measures have been suggested where each measure reflects different aspects of the surveillance method, e.g. the probability of detecting a true change within a certain period.

1.3. Notations and Specifications

We are monitoring a rare health event, the occurrence of a specific congenital malformation in a well-defined population. We are interested in discovering a sudden shift from the accepted 'normal' rate of this malformation to an increased rate.

The random process determining the state of the system is denoted $p(u)$, $0 < p(u) < 1$, $u = 1,2,3,\ldots$. We want to distinguish between the 'normal' rate of a congenital malformation p_0 and the increased rate p_1. A shift can occur at any time point τ during the surveillance period. Thus $p(u) = p_0$ for u=1, ..., τ-1 and $p(u) = p_1$ for u=τ, τ+1, ... In Section 4.5 τ is a random variable with a specified density. The incidence of a shift, $inc(t') = P\{t = t' | t \geq t'\}$, is assumed constant.

In other words, we view the surveillance situation as a random process with an accepted 'normal' rate of a congenital malformation but also with a risk of a 'catastrophe' present. If the lurking 'catastrophe' occurs it will manifest itself in a sudden shift to a higher level, i.e. a random process with an increased rate of the congenital malformation we are monitoring. Kennett and Pollak (1983) point out that modelling the change as a sudden shift is a simplification and that some kind of linear trend is a more realistic situation. Sveréus (1995) gives a detailed discussion of a surveillance situation where the increase is a linear trend. Wessman discusses a multiple surveillance situation where several processes are observed simultaneously.

Successive decisions must be taken and at decision point *s* the aim is to discriminate between the critical event *C(s)* and its complement *D(s)*. The critical event is $C(s) = \{t \leq s\} = \{p(s) = p_1\}$ the event that a shift occurs at *s* or

A(s) is the alarm set, i.e. a set of events with the property that when an observation belongs to *A* it is an indication that *C* occurs and a hypothesis stating a stable system is rejected. If an alarm is triggered the surveillance situation changes. Epidemiological research is needed to determine if the alarm is a false one or not and decisions on actions have to be taken. Thus only a first alarm is considered and the surveillance is regarded as being active in the sense of Frisén and deMaré (1991).

2. THE SETS METHOD

The Sets method, introduced by Chen (1978 and 1986), is a method especially developed for surveillance of rare health events. Distinguishing it from most other methods commonly used in this field.

With the Sets method the test is carried out each time an event under surveillance occurs, in our case each time a baby with a congenital malformation is born. The variable used is the period between two consecutive births of babies with a specific congenital malformation. This period can be expressed by the time interval between the occurrence of two consecutive events or as the number of healthy babies born between two consecutive births of malformed babies. Here the latter expression is used.

We then have a sequence of independent Bernoulli trials were π is the probability of each trial resulting in a malformed baby. Accordingly, the number of healthy babies born between two consecutive births of malformed babies is a geometrically distributed variable, X(j), where j is the j-th malformed baby born.

We define a threshold value, $T = \frac{k}{\pi_0}$ where k is a parameter of the Sets method. Furthermore, we define $p = P\{X(j) < T\}$ in general and

$$p_0 = P\{X(j) < T | \pi = \pi_0\}$$
$$p_1 = P\{X(j) < T | \pi = \pi_1\}$$

In order to give an alarm, n consecutive realisations of X(j) must be shorter than the prespecified threshold value T. The Sets method thus has two parameters: k specifies the threshold value and n specifies how far back in history we want to go.

If X(j) is assumed to be exponentially distributed, cf. Chen (1978), then

$$p_0 = 1 - e^{-k}$$
$$p_1 = 1 - e^{-\gamma k}$$

where $\gamma = \frac{\pi_1}{\pi_0}$, $\gamma > 1$.

For illustrations of results, the parameters of the Sets method are chosen as n=2 and k=*0.2287*. These values were used by Gallus (1986). The value of k determines the value of p_0 in accordance with the exponential assumption above. The examples are calculated for different values of p_1, thus defining the values of γ.

Chen (1978) presented a simple definition of an alarm, being n consecutive short intervals when we are at the decision point s

$$A(s) = \{[X(s-n+1) < T] \cap [X(s-n+2) < T] \cap \ldots \cap [X(s) < T]\}$$

Gallus(1986) presented a new definition of an alarm, suggesting that the n consecutive short intervals should <u>not</u> be preceded by a short interval, i.e. when s=n

$$A(s) = \{[X(s-n+1) < T] \cap [X(s-n+2) < T] \cap \ldots \cap [X(s) < T]\}$$

and when s>n

$$A(s) = \{[X(s-n) \geq T] \cap [X(s-n+1) < T] \cap \ldots \cap [X(s) < T]\}$$

The practical consequence of this definition is, of course, that a second alarm can only be triggered after n+1 events, i.e. $A(s) \cap A(s+i) = \varnothing,\ i \leq n$. The alarm definition used from now on is the definition presented by Gallus.

Since the surveillance situation is active, we are only interested in the first alarm. Hence we define <u>an active alarm set</u> as

$${}^{a}A(s) = A(s) \cap A^{c}_{s-1} \text{ where } A^{c}_{s-1} = A^{c}(n) \cap \ldots A^{c}(s-1),$$

i.e. the set of events giving an alarm at s but no alarm earlier.

3. MEASURES OF PERFORMANCE

The significance level α and the power of the test $1-\beta$, being the usual measures for evaluation of the performance of statistical tests, are inappropriate in the surveillance situation. This is due to the fact that these measures do not account for the time dimension present in the surveillance situation. Other measures taking this into consideration have been suggested in the literature, cf. Frisén (1992) and Åkermo (1994).

All measures depend on the run length distribution in some manner. Bear in mind that most systems of surveillance lead eventually to an alarm, whether a change has occurred or not. Here a run is defined as the period from the start of the surveillance until we get an alarm. Run length (RL) is defined as the length of this period, here being the number of events occurring until an alarm is triggered. The run length in terms of surveillance of congenital malformations is thus the number of infants born with a congenital malformation before an alarm is triggered for the first time.

3.1. Average Run Length

A common measure, developed in the field of quality control, is the Average Run Length (ARL) until an alarm is activated. ARL_0 is the average number of times an event, under surveillance, is observed until an alarm is triggered, when no change has occurred. ARL_1 is the average number of babies born with a congenital malformation until an alarm, in the case where the change occurred before the surveillance started. ARL is commonly used as an optimality criterion.

ARL is, in the context of surveillance of congenital malformations, useful as a crude measure of when in the surveillance period an alarm is triggered, i.e. how quick we get an alarm. The RL-distribution is a skew distribution. More detailed information requires other measures in addition to the ARL measure.

3.2. The Probability of a False Alarm

Several false alarm probabilities are described in the literature. Three of these discussed here are: the probability of a false alarm at the decision point s, the cumulative false alarm probability and the conditional false alarm probability.

The probability of a false alarm exactly at decision point s is

$$\begin{aligned}\alpha^*(s) &= P\{RL = s|D(s)\}\\ &= P\{{}^aA(s)|D(s)\}\\ &= P\{A(s)\cap A_{s-1}^c|D(s)\}\\ &= P\{\text{first alarm at decision point s}|\text{no change has occurred so far}\}\end{aligned}$$

This α-measure thus tells us how likely it is that when a infant with a congenital malformation is born we would conclude wrongly, for the first time, that the 'normal' rate of this malformation has indeed increased. It converges to zero when $s\to\infty$ for most methods, cf. Section 4.2, i.e. the longer we follow the process the smaller the probability of a <u>first</u> false alarm.

A second interesting measure is the cumulative false alarm probability.

$$\begin{aligned}\alpha_s &= P\{RL \le s|D_s\}\\ &= P\{A_s|D_s\}\\ &= P\{{}^aA_s|D(s)\}\\ &= P\{\text{alarm no later than at decision point s } | \text{ no change has occurred so far}\}\end{aligned}$$

where ${}^{a}A_s = {}^{a}A(1) \cup \ldots \cup {}^{a}A(s) = \mathrm{A}(1) \cup \ldots \cup \mathrm{A}(\mathrm{s}) = \mathrm{A}_{\mathrm{s}}$
and $D_s = D(1) \cap \ldots \cap D(s)$.

Note that $D_s = D(s)$ since $D(s) = \{\tau > s\}, s = 1,2,\ldots,$ is a non increasing sequence of sets, cf. Frisén and deMaré (1991).

This α-measure does, in contrast to the first one, tell us how likely it is that at the birth of the *s*-th baby, with a congenital malformation, at least one false alarm is triggered then or has been triggered before. For methods like the Sets method, it can be expected to converge to 1 when *s* increases, cf. Section 4.3.

A third similar measure is called the conditional false alarm probability. This is the probability of getting a false alarm at *s* conditioned on not having had any false alarms earlier.

$$\begin{aligned}\alpha(s) &= P\{RL = s \mid RL > s-1, D(s)\} \\ &= P\{A(s) \mid A^{c}_{s-1}, D(s)\} \\ &= P\{\text{alarm at decision point s} \mid \text{no alarm earlier, no change has occurred so far}\}\end{aligned}$$

In other words, when *s-1* babies with a birth defect have been born and no false alarms have been triggered so far, how likely it is to get a false alarm the next time a baby with a birth defect is born. This α-measure is the probability of getting a false alarm at decision point *s* when we are at decision point *s-1* and have not got an alarm.

These three preceding α-measures illustrate different aspects of time vs. false alarm probability. The next two measures focus instead on the power of the method.

3.3. The Probability of a Successful Detection

A measure reflecting the power of the surveillance method is the probability of a successful detection (*PSD*). This is the probability of detecting a true change occurring at t', i.e. $\tau=t'$, within a certain interval $t' \leq s < t'+d$, on condition that no alarm has been triggered before t'. In other words, if a 'normal' rate of a congenital malformation suddenly increases at t' how likely is it that we discover the change before d more malformed babies are born, given we have not had any false alarms before t'.

$$
\begin{aligned}
PSD\{t',d,\gamma\} &= P\{RL \leq t'+d-1 | RL > t'-1, \tau = t'\} \\
&= P\{A_{t'+d-1} | A^{c}_{t'-1}, \tau = t'\} \\
&= P\{\text{alarm at least at } t'+d-1 \,|\text{no alarm before } t', \text{a shift occurred at } t'\}
\end{aligned}
$$

We would want this measure to be high since the 'cost' of d more malformed babies is hard to accept. For a constant t' and γ, the greater d gets the easier it is to discover the increase before d more babies are born, i.e. *PSD* converges to 1 for increasing d.

3.4. The Predictive Value of an Alarm

The predictive value of an alarm is the relative frequency of motivated alarms among all alarms at a certain decision point. In other words this measure tells us how likely it is that a true change has occurred when we have had an alarm.

$$
\begin{aligned}
PV\{t'',\gamma,inc\} &= P\{\tau \le t'' | RL = t''\} \\
&= P\{C(t'') |^{a} A(t'')\} \\
&= P(\text{a shift has occurred at t'' or earlier} \mid \text{first alarm at t''})
\end{aligned}
$$

We would, of course, want a high predictive value, i.e. we would like to be able to rely on that an alarm indicates a change in the 'normal' rate of a congenital malformation under surveillance.

4. RESULTS

4.1. The Probability of an Alarm

Here we give results on the probability of an alarm which will be used in subsequent sections.

According to the definition of an alarm by the Sets method, presented in Section 2, the proper sequence of events giving an alarm at the decision point $s>n$ is

$$A(s) = \{[X(s-n) \geq T] \cap [X(s-n+1) < T] \cap \ldots \cap [X(s) < T]\}$$

where $T = \frac{k}{\pi_0}$, n and k are the parameters of the Sets method.

Hence n=2 gives $A(s) = \{[X(s-2) \geq T] \cap [X(s-1) < T] \cap [X(s) < T]\}$ for $s>n$.

Gallus (1986) and Kenett and Pollak (1983) discuss the probability of <u>a first alarm</u> at a certain decision point s, $P\{{}^{a}A(s)\} = P\{A(s) \cap A^{c}_{s-1}\}$. Feller (1950) discusses the same probability in the context of the theory of recurrent events and success runs.

When $s<n$ no alarm can be triggered, but when $s=n$ we have

$$P\{{}^{a}A(n)\} = P\{A(n)\} = P\{[X(1) < T] \cap \ldots \cap [X(n) < T]\} = p^{n}$$

where p is defined in Section 2.

Likewise, when $n<s\leq 2n$, the probability of a first alarm exactly at the decision point s is

$$\begin{aligned} P\{{}^{a}A(s)\} &= P\{A(s)\} \\ &= P\{[X(s-n) \geq T] \cap [X(s-n+1) < T] \cap \ldots \cap [X(s) < T]\} \\ &= (1-p)p^{n} \end{aligned}$$

When $s>2n$ a more complicated situation arises since several alarm sequences could have occurred before the decision point s.

Using the following facts

$A(s) \cap A(s+i) = \varnothing$ when $i \le n$,

$A(s)$ and $A(s+i)$ are independent when $i > n$.

we get

$$\begin{aligned} P\{{}^{a}A(s)\} &= P\{A(s) \cap A_{s-1}^{c}\} = P\{A(s) \cap A_{s-n-1}^{c}\} \\ &= P\{A(s)\}P\{A_{s-n-1}^{c}\} \\ &= P\{A(s)\}\left[1 - P\{{}^{a}A_{s-n-1}\}\right] \\ &= P\{A(s)\}\left[1 - \sum_{j=n}^{s-n-1} P\{{}^{a}A(j)\}\right] \end{aligned}$$

In order to simplify the notations when calculating the probability of a successful detection and the predictive value of an alarm, discussed in Section 4.5 and 4.6 respectively, we introduce a notation for the probability of a first alarm at t'' when a change has occurred at t'

$$P\{{}^{a}A(t'') | \tau = t'\} = \omega(t'' | t')$$

Observe that when $t''<t'$ we have the probability of a first false alarm at t'' discussed in Section 4.2.

4.2. The Probability of a False Alarm at each Decision Point

We have

$$\begin{aligned}\alpha^*(s) &= P\{{}^aA(s)|D(s)\} \\ &= P\{A(s)\cap A_{s-1}^c|D(s)\} \\ &= P\{\text{first alarm at decision point s}|\text{no change has occurred so far}\}\end{aligned}$$

By definition, an alarm cannot occur before n observations, i.e. $\alpha^*(s) = 0$ when $s=1, \ldots, n-1$.

When exactly n observations are available, i.e. when $s=n$, we get $\alpha^*(n) = p_0^n$.

As in Section 4.1, only one alarm sequence can occur when $n < s \leq 2n$

$$\alpha^*(s) = P\{{}^aA(s)|D(s)\} = P\{A(s)|D(s)\} = p_0^n(1-p_0).$$

If $s>2n$, using the relation in Section 4.1, we have

$$\begin{aligned}\alpha^*(s) &= P\{{}^aA(s)|D(s)\} \\ &= P\{A(s)|D(s)\}\left[1-\sum_{j=n}^{s-n-1} P\{{}^aA(j)|D(s)\}\right] \\ &= \alpha^*(n+1)\left[1-\sum_{j=n}^{s-n-1}\alpha^*(j)\right]\end{aligned}$$

According to Feller (1950), we have $\sum_{j=n}^{\infty}\alpha^*(j) = 1$. In view of the results in Section 4.3 where $\alpha_s = \sum_{j=n}^{s}\alpha^*(j)$ when $s>2n$, we can then calculate the limit value $\lim_{s\to\infty}\alpha^*(s) = \lim_{s\to\infty}[\alpha_s - \alpha_{s-1}] = 0$.

Figure 1 confirms these results and demonstrates the characteristic look of the plot of the false alarm probability at each decision point $\alpha^*(s)$ vs. the decision point s. Note especially the distinct peek at $s=n$.

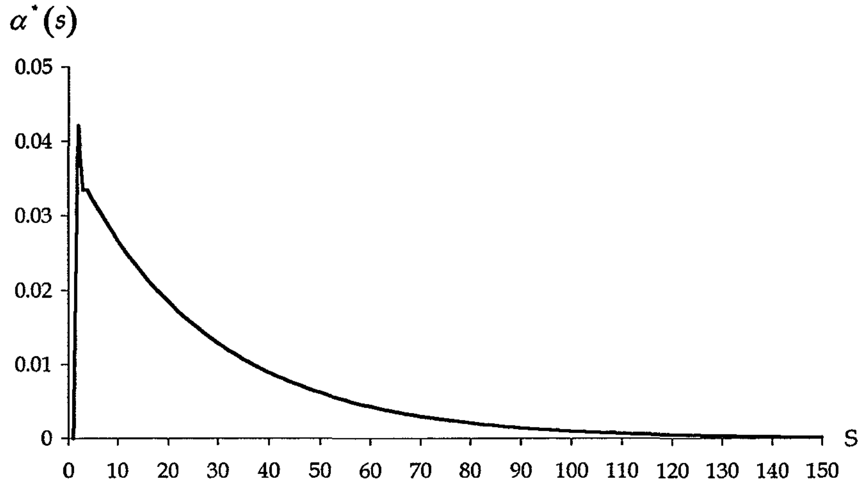

Figure 1 The probability of a first false alarm $\alpha^*(s)$ at decision point s for the Sets method, n=2 and k=0.2287.

4.3. Cumulative False Alarm Probability

We have

$$\alpha_s = P\{A_s|D_s\}$$
$$= P\{\text{alarm no later than at decision point s}|\text{no change has occurred so far}\}$$

As before, an alarm cannot occur before n observations have been achieved, i.e. $\alpha_s = 0$ when $s=1, \ldots, n-1$.

If $s \geq n$ the cumulative false alarm probability is

$$\begin{aligned}\alpha_s &= P\{A_s | D(s)\} \\ &= P\{{}^aA_s | D(s)\} \\ &= \sum_{j=n}^{s} P\{{}^aA(j) | D(j)\} \\ &= \sum_{j=n}^{s} \alpha^*(j)\end{aligned}$$

using the fact that ${}^aA(j') \cap {}^aA(j'') = \varnothing,\ j' \neq j''$.

We know from Feller (1950) that $\lim_{s\to\infty} \alpha_s = \sum_{j=n}^{\infty} \alpha^*(j) = 1$. Figure 2 confirms these results and demonstrates the characteristic look of the plot of the cumulative false alarm probability α_s vs. the decision point s.

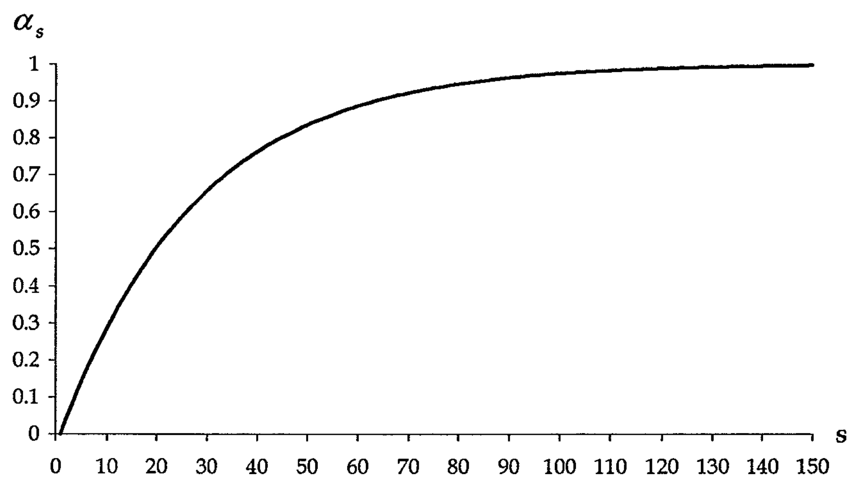

Figure 2 The cumulative probability of a false alarm α_s at decision point s for the Sets method, n=2 and k=0.2287.

4.4. The Conditional Probability of a False Alarm at each Decision Point

We have

$$\alpha(s) = P\{A(s) | A^c_{s-1}, D(s)\}$$
$$= P\{\text{alarm at decision point s} | \text{no alarm earlier, no change has occurred}\}$$

By definition an alarm cannot occur before n observations have been achieved, i.e. $\alpha(s) = 0$ when $s=1, \ldots, n-1$. When $s \geq n$ the conditional false alarm probability is

$$\alpha(s) = P\{A(s) | A^c_{s-1}, D(s)\}$$
$$= \frac{P\{A(s) \cap A^c_{s-1} | D(s)\}}{P\{A^c_{s-1} | D(s)\}}$$
$$= \frac{\alpha^*(s)}{(1-\alpha_{s-1})}$$

since $A^c_{s-1} = {}^a A^c_{s-1}$.

When $\lim_{s\to\infty} \frac{\alpha^*(s-1)}{\alpha^*(s)} = \lambda$ it follows that $\lim_{s\to\infty} \frac{\alpha^*(s-n-1)}{\alpha^*(s-1)} = \lambda^n$ since

$$\frac{\alpha^*(s-n-1)}{\alpha^*(s-1)} = \frac{\alpha^*(s-n-1)}{\alpha^*(s-n)} \frac{\alpha^*(s-n)}{\alpha^*(s-n+1)} \cdots \frac{\alpha^*(s-2)}{\alpha^*(s-1)}.$$

A reformulation of the expression for $\alpha^*(s)$, cf. Section 4.2, gives for $s \geq n$

$$\alpha^*(s) = \alpha^*(s-1) - \alpha^*(n+1)\alpha^*(s-n-1)$$
$$\Leftrightarrow$$
$$\frac{\alpha^*(s)}{\alpha^*(s-1)} = 1 - \alpha^*(n+1)\frac{\alpha^*(s-n-1)}{\alpha^*(s-1)}$$

thus giving an equation determining the value of λ

$$\frac{1}{\lambda} = 1 - \alpha^*(n+1)\lambda^n.$$

Note that λ is then a constant depending only on n and k.

Furthermore the expression for $\alpha(s)$ can be rewritten in the following manner

$$\alpha(s) = \frac{\alpha^*(s)}{(1-\alpha_{s-1})} = \alpha^*(n+1)\frac{\alpha^*(s)}{\alpha^*(n+s)}$$

giving

$$\lim_{s\to\infty} \alpha(s) = \alpha^*(n+1)\,\lambda^n = 1 - \frac{1}{\lambda}.$$

Figure 3 confirms these results and demonstrates the characteristic look of the plot of the conditional false alarm probability $\alpha(s)$ vs. the decision point s. Note especially the distinct peek at s=n.

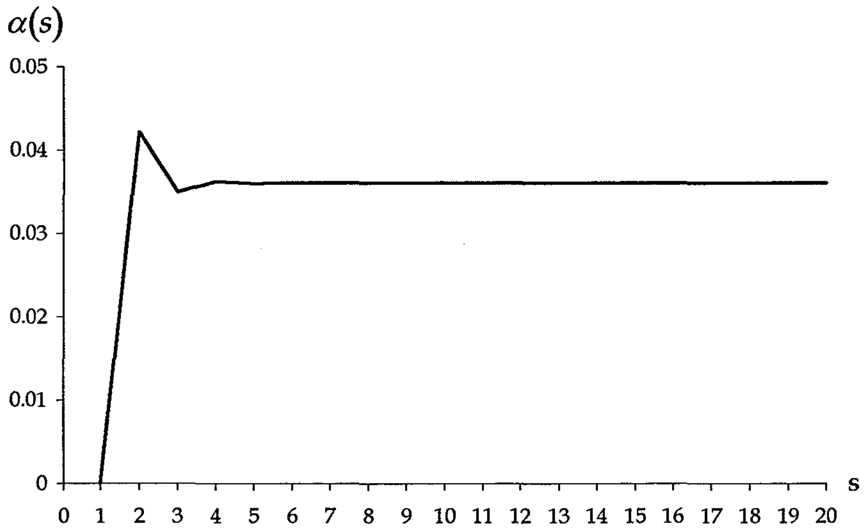

Figure 3 The conditional probability of a false alarm $\alpha(s)$ at each decision point s for the Sets method, n=2 and k=0.2287. $\lim_{s\to\infty} \alpha(s) = 0.036$.

4.5. The Probability of a Successful Detection

The probability of a successful detection (*PSD*) is the probability to discover a true shift within *d* units after the change has occurred given that no alarm has been triggered earlier. *PSD* is dependent on the size of the shift γ, the length of the detection interval *d* and the time point of the change *t'*.

We have

$$PSD\{t',d,g\} = P\{A_{t'+d-1} | A^c_{t'-1}, t=t'\}$$
$$= \frac{P\left\{\bigcup_{s=t'}^{t'+d-1} {}^aA(s) | t=t'\right\}}{P\{A^c_{t'-1} | t=t'\}}$$
$$= \frac{\sum_{s=t'}^{t'+d-1} P\{{}^aA(s) | t=t'\}}{1-a_{t'-1}}$$
$$= \frac{\sum_{s=t'}^{t'+d-1} w(s|t')}{1-a_{t'-1}}$$

since $A^c_{s-1} = {}^aA^c_{s-1}$ and using the notation $P\{{}^aA(s) | t=t'\} = w(s|t')$ from Section 4.1.

Similar to Section 4.1, when *t'*>2*n* the above relation can be written as

$$PSD\{t',d,g | t'>2n\} = \frac{\sum_{s=t'}^{t'+d-1}\left[P\{A(s)|t=t'\}\left(1-\sum_{j=n}^{s-n-1} w(j|t')\right)\right]}{1-a_{t'-1}}$$

Two illustrative special cases are subjects of a closer study: $PSD\{t'=1,d,g\}$ and $PSD\{t',d=1,g\}$.

The first special case studied is where *t'*=1. This implies that the malformation rate is at the increased level when surveillance starts. The process is thereby in process is thereby in the same state, the increased level, at all decision points including the present one. This can be compared with the cumulative false alarm probability where the process has been in the 'normal' state at all decision points.

When $d<n$ we have $PSD\{t'=1,d,\gamma\}=0$ but when $d\geq n$ we get

$$PSD\{t'=1,d,\gamma\}=\sum_{s=n}^{d}\omega(s|1).$$

Note that we sum from n since by definition $\omega(s|1)=0$ when $s<n$.

In similarity to previous measures, when $d=n$ we get

$$PSD\{t'=1,d,\gamma\}=\omega(n|1)=p_1^n.$$

When $n<d\leq 2n$ we get

$$\begin{aligned}PSD\{t'=1,d,\gamma\}&=\omega(n|1)+\sum_{s=n+1}^{d}\omega(s|1)\\&=p_1^n+(d-n)(1-p_1)p_1^n\end{aligned}$$

Finally when $d>2n$ we get

$$\begin{aligned}PSD\{t'=1,d,\gamma\}&=\omega(n|1)+\sum_{s=n+1}^{2n}\omega(s|1)+\sum_{s=2n+1}^{d}\left[P\{A(s)|\tau=1\}\left(1-\sum_{j=n}^{s-n-1}\omega(j|1)\right)\right]\\&=p_1^n+n(1-p_1)p_1^n+(1-p_1)p_1^n\sum_{s=2n+1}^{d}\left(1-\sum_{j=n}^{s-n-1}\omega(j|1)\right)\end{aligned}$$

Figure 4 illustrates the characteristics of the plot of the $PSD\{t'=1,d,\gamma\}$ vs. d. Note the steep increase from zero for $d<2$ to 0.81 for $d=2$. This must be seen in light of the parameter values chosen for n and k. Similar to α_s, the cumulative false alarm probability, in Section 4.3, $\lim_{d\to\infty}PSD\{t'=1,d,\gamma\}=1$. Note that

$PSD\{t'=1,d,\gamma\}$ converges even faster than α_s since the probability of an alarm is much higher in the increased state.

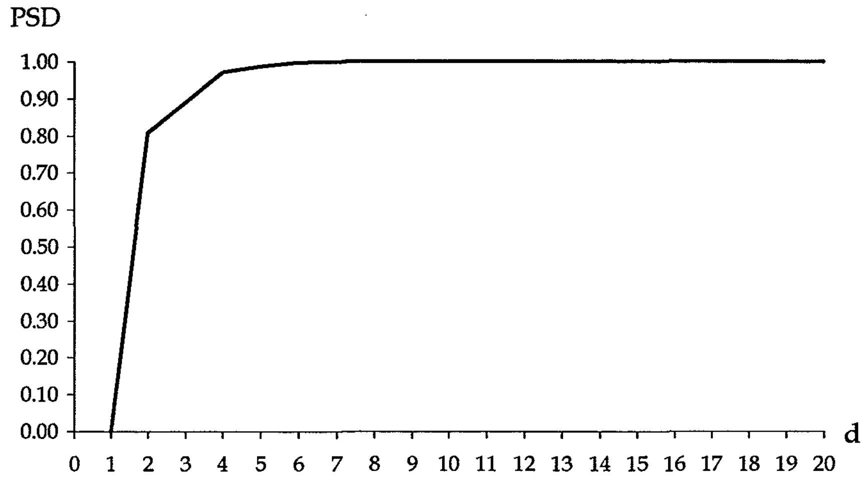

Figure 4 A plot of the probability of a successful detection $PSD\{t'=1,d,\gamma\}$ vs. d for the Sets method, n=2 and k=0.2297. This is the special case of t'=1, i.e. a shift, γ=10, occurred at the beginning.

The second special case, where d=1, demonstrates other important aspects of the *PSD* measure. This is the probability of discovering the change at the same time it occurs given that we have had no false alarms earlier. Note the parallel to the conditional false alarm probability α(s) in Section 4.4.

The following relation holds for t'=*1, 2,…* .

$$PSD\{t',d=1,\gamma\}=\frac{\omega(t'|t')}{1-\alpha_{t'-1}}$$

When $t'<n$ we have $PSD\{t',d=1,\gamma\}=0$, but when $t'=n$ we get

$$PSD\{t',d=1,\gamma\}=P\{A(n)|\tau=n\}=p_0^{n-1}p_1.$$

When $n<t'\leq 2n$ we get $PSD\{t',d=1,\gamma\}=\dfrac{\omega(t'|t')}{1-\alpha_{t'-1}}=\dfrac{(1-p_0)p_0^{n-1}p_1}{1-\alpha_{t'-1}}$

and finally $t'>2n$ gives

$$PSD\{t',d=1,\gamma\}=\frac{P\{A(t')|\tau=t'\}\left(1-\sum_{j=n}^{t'-n-1}\omega(j|t')\right)}{1-\alpha_{t'-1}}=(1-p_0)p_0^{n-1}p_1\frac{(1-\alpha_{t'-n-1})}{(1-\alpha_{t'-1})}$$

In order to examine $\lim_{t'\to\infty}PSD\{t',d=1,\gamma\}$, we rewrite the ratio in the above expression

$$\frac{(1-\alpha_{t'-1})}{(1-\alpha_{t'-n-1})}=\frac{1-\alpha_{t'-n-1}-\sum_{j=t'-n}^{t'-1}\alpha^*(j)}{1-\alpha_{t'-n-1}}=1-\frac{\sum_{j=t'-n}^{t'-1}\alpha^*(j)}{1-\alpha_{t'-n-1}}=1-\frac{\alpha^*(n+1)\sum_{j=t'-n}^{t'-1}\alpha^*(j)}{\alpha^*(t')}$$

using the relation from Section 4.3 where $\alpha^*(s)=\alpha^*(n+1)[1-\alpha_{s-n-1}]$ for $s>2n$.

When $\lim_{s\to\infty}\dfrac{\alpha^*(s-1)}{\alpha^*(s)}=\lambda$, see Section 4.4, the case of n=2 yields

$$\frac{(1-\alpha_{t'-1})}{(1-\alpha_{t'-3})}=1-\alpha^*(3)\left[\frac{\alpha^*(t'-2)}{\alpha^*(t')}+\frac{\alpha^*(t'-1)}{\alpha^*(t')}\right]\underset{t'\to\infty}{\to}1-\alpha^*(3)\left[\lambda^2+\lambda\right].$$

Which gives

$$\lim_{t'\to\infty}PSD\{t',d=1,\gamma\}=(1-p_0)p_0^{n-1}p_1\frac{1}{1-(1-p_0)p_0^2\left[\lambda^2+\lambda\right]}$$

This limit value is 0.16, see Figure 5, for k=0.2287 and γ=10. Figure 5 illustrates the characteristics of the plot of $PSD\{t',d=1,\gamma\}$ vs. t'. Note the similarity to Figure 3.

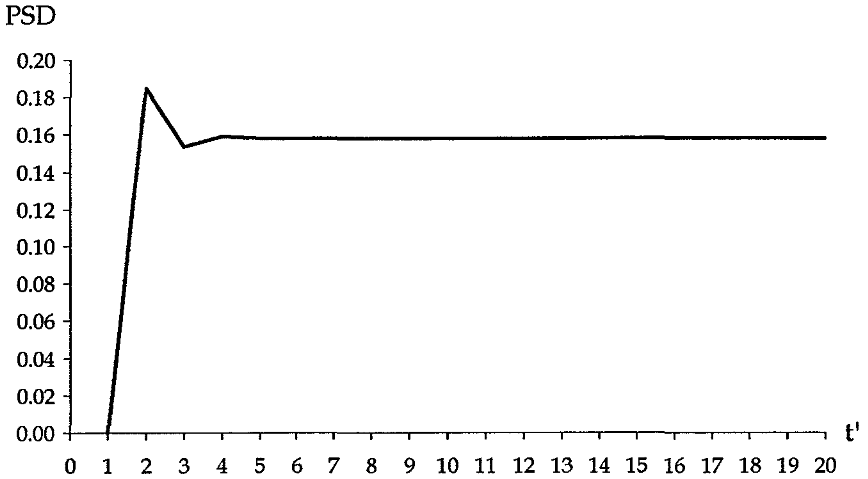

Figure 5 A plot of the probability of a successful detection $PSD\{t', d = 1, \gamma\}$ vs. t' for the sets method, n=2, k=0.2287 and γ=10. This is a special case of d=1, i.e. we want to discover the change as soon as it occurs.
$\lim_{t' \to \infty} PSD\{t', d = 1, \gamma\} = 0.16$

4.6. The Predictive Value of an Alarm

The predictive value of an alarm (*PV*) is the probability of a change being present at occasion t'' given that an alarm has been triggered at occasion t''. It measures how much an alarm should be trusted.

PV is dependent on the alarm occasion t'', the size of the shift γ and the incidence *inc*. The changing point τ is regarded as random. The incidence of a change is $inc(t') = P\{\tau = t' | \tau \geq t'\}$. We will assume that the incidence is constant *inc*, τ thus has a geometric distribution.

Bayes's rule gives us

$$PV\{t'',\gamma,inc\} = P\{C(t'')|{}^{a}A(t'')\}$$
$$= \frac{P\{C(t'')\}P\{{}^{a}A(t'')|C(t'')\}}{P\{D(t'')\}P\{{}^{a}A(t'')|D(t'')\} + P\{C(t'')\}P\{{}^{a}A(t'')|C(t'')\}}$$
$$= \frac{\sum_{t'=1}^{t''} inc(1-inc)^{t'-1}\omega(t''|t')}{(1-inc)^{t''}\alpha^{*}(t'') + \sum_{t'=1}^{t''} inc(1-inc)^{t'-1}\omega(t''|t')}$$

using the notation $P\{{}^{a}A(t'')|\tau = t'\} = \omega(t''|t')$ from Section 4.1.

Exact calculations of PV are tedious but Figure 6 demonstrates simple cases of PV for the Sets method with n=2 and k=0.2287, where γ=10 and *inc* is 0.001, 0.01 and 0.1 respectively. In Figure 6 it can be seen that the incidence *inc=0.001* gives extremely low PV at a seemingly constant level. This constant level is really a slight monotone increase. When *inc=0.01* the level of the PV measure increases and an extreme increase can be seen with *inc=0.1* We therefore conclude that the incidence is of crucial importance for the level of the PV measure.

The incidence is the probability of the 'catastrophe' occurring at a certain decision point. Since the Sets method involves each birth of a malformed baby as a decision point, the value of the incidence must thus be interpreted in the context of malformed babies. For example, *inc=0.1* means that we believe the risk of a 'catastrophe' leading to a γ-high increase in the 'normal' rate of a malformation to be ten percent. In order to lend ten percent some meaning this can be rephrased in the following way: we expect a 'catastrophe' leading to a γ-high increase in the 'normal' rate of a malformation at every tenth birth of baby with that malformation. The meaning of the incidence in calendar time is therefore dependent on the rarity of the malformation under surveillance and the population size.

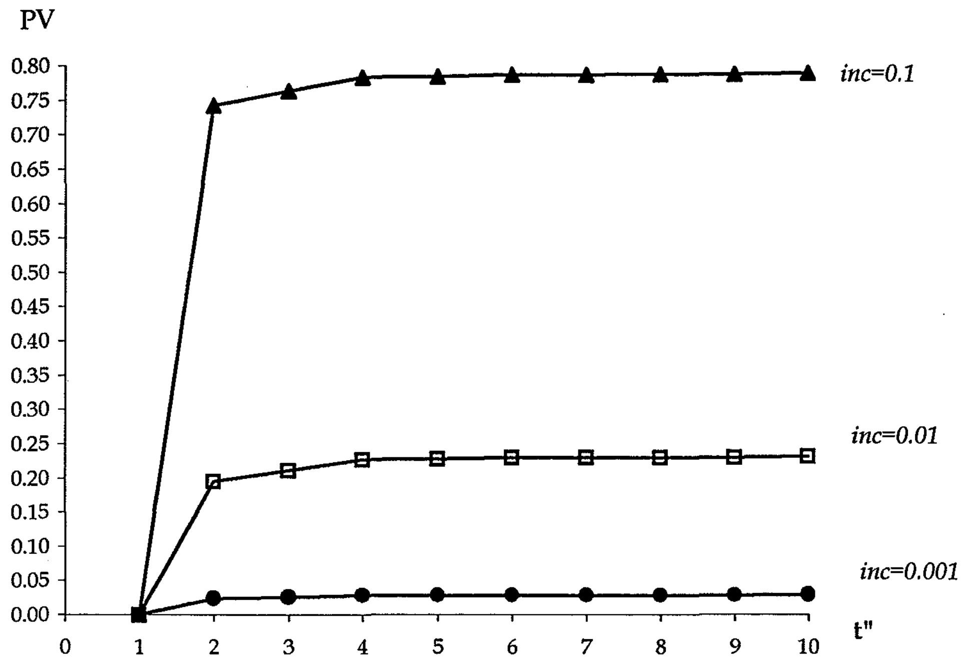

Figure 6 A plot of the predictive value of an alarm $PV\{t'', \gamma, inc\}$ vs. t'' for the Sets method, n=2 and k=0.2297. The size of the shift γ is 10 and the incidence *inc* is 0.001, 0.01 and 0.1 respectively.

Figure 7 shows PV for *inc=0.1* with the size of the shift γ equal to 2, 6 and 10, respectively. It is seen that changing the size of the shift does not affect the level of the PV measure in the same exceptional manner. Instead a change in pattern is seen where the higher values, γ=6 and γ=10, achieve their level faster than the low level,γ=2. The difference is most obvious when γ=10 and γ=2 are compared.

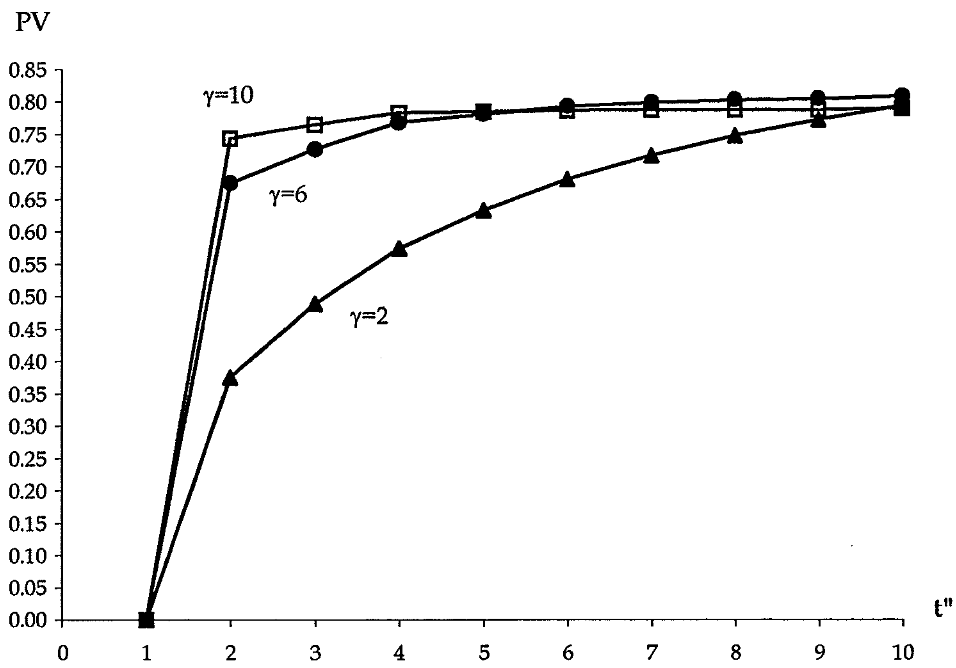

Figure 7 A plot of the predictive value of an alarm $PV\{t'',\gamma,inc\}$ vs. t'' for the Sets method, n=2 and k=0.2287. The incidence of a shift is inc=0.1 and the size of the shift γ is 2, 6 and 10 respectively.

5. CONCLUDING REMARKS

One of the objectives of this paper was to derive measures for evaluation of the performance of the Sets method. The traditional evaluation measure used is ARL. Other measures reflect different aspects of the Sets method and might eventually be better suited depending on the situation at hand. Explicit expressions were obtained for the three false alarm measures: the probability of a false alarm at each decision point $a^*(s)$, the cumulative false alarm probability a_s and the conditional false alarm probability $a(s)$. The probability of a successful detection (*PSD*) as well as the predictive value of an alarm (*PV*) resulted in general expressions and their properties were discussed in the contexts of special cases and examples. For all measures the results were illustrated using the same example namely the Sets method with the parameters chosen as n=2 and k=*0.2287* and the size of the shift γ=10.

The probability of a false alarm at each decision point $a^*(s)$ peaks at s=n, remains constant for $n<s\leq 2n$ and converges towards zero as s grows, see Figure 1 for example. The relation between the cumulative false alarm probability a_s and $a^*(s)$, and the early peak of $a^*(s)$, leads to a very high cumulative false alarm even for moderately high $a^*(s)$ levels. This is demonstrated in Figures 1 and 2 where a_s is 0.11 already at s=4. The conditional probability of an alarm $a(s)$ peeks at s=n, in the same way as $a^*(n)$, but converges very fast towards a limit value which can be calculated for relevant values of n and k, see Figure 3 for example.

The probability of a successful detection *PSD* measures the power of the method. The first special case $PSD\{t'=1,d,g\}$ where the change has occurred at the beginning is essentially a cumulative power measure, analogous to the cumulative false alarm probability. In Figure 4 it converges towards one in the same way as the cumulative false alarm but at a faster rate. Figure 4 shows a level of 0.81 already at d=2 in the example. The second special case $PSD\{t', d=1, \gamma\}$ is comparable to the conditional false alarm $\alpha(s)$ since it measures how likely it is that we discover a true change at the same time it occurs, conditioned on the fact that we have had no alarm earlier. $PSD\{t', d=1, \gamma\}$ peaks at $t'=n$ and converges swiftly towards a limit value. In the example illustrated the maximum value of $PSD\{t', d=1, \gamma\}$ is 0.18 and the limit value, reached already at $t'=9$, is 0.16. This is a lower value than would be preferred especially since the size of the shift is considerable.

Simulation results indicate that the *PSD* measure converges with increasing t' even when $d>1$, demonstrating the same pattern as $PSD\{t', d=1, \gamma\}$. This can be interpreted as follows: *'after an initial period, it is of no importance for the PSD measure where the change occurs in the surveillance period'*. A plausible explanation to this could be the fact that any decision upon an alarm is based on n observations, regardless of how long the surveillance has been in progress.

The predictive value of an alarm measures the degree of trust we should have in an alarm and seems to be extremely dependent on how likely we consider that a 'catastrophe' would happen, i.e. the incidence. It is also dependent on the size of the shift γ although this dependency does not appear to be as distinct.

This knowledge of how the different measures interact is very valuable since it opens new possibilities of how to vary the parameters of the Sets method in order to achieve certain qualities in a specific surveillance situation.

This work should be considered as an appetiser and, at least in the authors mind, gives rise to several ideas of how the work could be proceeded.

ACKNOWLEDGEMENTS

This work has been a part of a project supported by the Swedish Council for Research in the Humanities and Social Sciences. I would like to thank my supervisor Marianne Frisén for her guidance and constant encouragement. I also thank my company Astra Hässle AB for the financial support I have recieved. I further thank my colleagues at the Department of Statistics, University of Göteborg as well as my colleagues at Astra Hässle AB for contributing with constructive discussions and companionship. I am grateful to Robert Johnson for valuable comments on the draft manuscript.

Källen, B. (1989). Population Surveillance of Congenital Malformations. Possibilities and Limitations. *Acta Paediatr Scand.* **78**, 657-663.

Lie, R.T., Vollset, S.T., Botting, B., 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?. *Int. Journal of Risk & Safety in Medicine.* **2**, 289-300.

Radaelli, R., Gallus, G. (1989). On Detection of a Change in the Dynamics of Rare Health Events. *Commun. Statist. -Theory Meth.* **18**, 579-590.

Radaelli, G. (1992). Using the Cuscore technique in the surveillance of rare health events. *Journal of Applied Statistics.* **19**, 75-81.

Shore, D.L., Quade, D. (1989). A Surveillance System Based on a Short Memory Scheme. *Statistics in Medicine.* **8**, 311-322.

Simpkin, J.M., Downham, D.Y. (1988). Testing for a Change-point in Registry Data with an Example on Hypospadias. *Statistics in Medicine.* **7**, 387-393.

Sitter, R.R., Hanrahan, L.P., Demets, D., Anderson, H.A. (1990). A Monitoring System to Detect Increased Rates of Cancer Incidence. *American Journal of Epidemiology.* **132**, 123-130.

Socialstyrelsen, Folkhälsoenheten. (1990). Det svenska Missbildningsregistrets verksamhet.

Sveréus, A. (1995). Statistical methods for postmarketing surveillance. Detection of successive changes. *Research Report 1995:2.* Department of Statistics, Göteborg.

Åkermo, G. (1994). On performance of methods for statistical surveillance. *Research Report 1994:7.* Department of Statistics, Göteborg.

Weatherall, J.A.C. (1988). Surveillance of congenital malformations and birth defects. In: Eylenbosch, W.J., Noah, N.D., eds. *Surveillance in Health and Disease,* Oxford. 101-114.

Wessman, P. A procedure for multivariate surveillance.(Unpublished).

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.

1994:1	Frisén, M.	Statistical surveillance of business cycles.
1994:2	Frisén, M.	Characterization of methods for surveillance by optimality.
1994:3	Frisén, M. & Cassel, C.	Visual evaluation of statistical surveillance.
1994:4	Ekman, C.	A comparison of two designs for estimating a second order surface with a known maximum.
1994:5	Palaszewski,B.	Comparing power and multiple significance level for step up and step fown multiple test procedures for correlated estimates.
1994:6	Åkermo, G.	Constant predictive value of an alarm.
1994:7	Åkermo, G.	On performance of methods for statistical surveillance.
1994:8	Palaszewski, B.	An abstract bootstrap based step down test procedure for testing correlated coefficients in linear models.
1994:9	Ekman, C.	Saturated designs for second order models.
1994:10	Ekman, C.	A note on rotatability.