Research Report
Statistical Research Unit
Göteborg University
Sweden

A monitoring system for detecting starts and declines of influenza epidemics

Eva Andersson

Mailing address:	Fax	Phone	Home Page:
Statistical Research Unit	Nat: 031-773 12 74	Nat: 031-773 10 00	http://www.stat.gu.se/stat
P.O. Box 660	Int: +46 31 773 12 74	Int: +46 31 773 10 00	
SE 405 30 Göteborg			
Sweden			

A Monitoring System for Detecting Starts and Declines of Influenza Epidemics

Eva Andersson

Statistical Research Unit
Göteborg University

Abstract

The aim is to detect an influenza outbreak as soon as possible. Data are weekly reports of number of patients showing influenza-like symptoms. At each additional observation we decide whether a change has occurred or not.

The methodology of statistical surveillance is used to construct an outbreak detection system. The report also demonstrates measures that reflect timeliness, such as the probability of successful detection within a specified time and the predictive value at different time points.

A new non-parametric approach is used. The cycles are estimated using only monotonicity restrictions. Also different approaches regarding the intensity of the change-process are compared. The pros and cons of using an empirical intensity are evaluated.

The time to an alarm is investigated, both for false and motivated alarms. When setting the alarm limit, there is a trade-off between the false alarms and the delay of motivated alarms. In the evaluation it must be considered that the performance of the system depends on when the outbreak begins, in relation to the start of the surveillance. If the outbreak begins at the same time as the surveillance is started, then the probability of early detection is 0.04, but if the surveillance is started 12 time points before the outbreak the detection probability is 0.43.

The non-parametric approach avoids miss-specifications of the base line. Even a „normal" miss-specification results in serious delay. Another drawback is that alarms at late time points have low predictive value.

An empirical prior for the intensity works well when the actual outbreak time agrees with the prior. But when the outbreak occurs „earlier than expected", the alarms are seriously delayed. A non-informative prior, however, works well.

1. Introduction

In on-going cyclical processes, for example influenza cycles, it is often of interest to detect a turning point (an influenza outbreak) as soon as possible after it has occurred. Many methods have been proposed for detection of turns, see (*1*) and (*2*).

An early indicator is the number of patients showing influenza-like symptoms. It can be an advantage to use this indicator instead of the number of confirmed cases (after a laboratory analysis), since the latter variable might take longer. A surveillance system is constructed with the aim to detect an outbreak at an early stage.

One aim for monitoring influenza activity is to, as soon as possible, spread the information, since an outbreak can put a heavy burden on the medical service. It is therefore important to identify the start of an influenza epidemic. According to the Swedish Institute for Infectious Disease Control, a reporting system can consist of weekly reports of the number of possible cases (patients showing influenza-like symptoms), combined with number of confirmed cases (after samples have been analyzed in a laboratory). In surveillance, repeated decisions are made: each week, at each new additional observation, we decide whether an outbreak has started or not. There are several well known methods of surveillance, such as the CUSUM method and the EWMA method. A scan statistic is suggested in (*3*) and (*4*). A review of surveillance in public health is given in (*5*).

In an alarm system there is always the risk of false alarms. We want an alarm system with few false alarms and high detection probability for an actual outbreak, but if the false alarm probability is small, then it also takes longer to detect an actual outbreak. When setting the alarm limit, there is a trade-off between the false alarms and the delay of motivated alarms.

Timeliness is very important: when we compare methods we must know how long it takes before each method indicates that an outbreak has occurred. Measures that reflect timeliness are probability of successful detection within a specified time limit and the predictive value of an alarm at a specific time. These measures are further discussed by Frisén in (*6*).

2. Objectives

The methodology of statistical surveillance is used to construct a system for outbreak detection. Generally, the aim of a surveillance system can be described as discriminating between the in-control process and the out-of-control process. Ideally, we want few alarms when the process is in-control and we want quick indication when the process has gone out-of-control. A system based on the likelihood ratio between in-control process and out-of-control process is optimal, see (*7*). The likelihood ratio has been shown to be equivalent to the posterior probability, used in many Hidden Markov model approaches (HMM). In (*8*) a surveillance system using an HMM is discussed.

In this report we demonstrate measures that reflect timeliness, e.g the probability of successful detection within a specified time and the predictive value at different time points. The evaluation measures often depend on the time of the alarm or the time when the outbreak begins, in relation to the start of the surveillance.

3. Methods

3.1 Non-parametric surveillance

As mentioned above, the likelihood ratio (LR) is optimal. The LR method gives the shortest expected delay for a given false alarm probability. The likelihood for the in-control process (base line, event D) and the out-of-control process (outbreak, event C) are compared. The LR method at time s is defined as

$$LR(s) = \frac{f(x_s|C) \cdot P(C)}{f(x_s|D) \cdot P(D)} > k_s,$$

where $f(x_s|m)$ is the likelihood for $x_s = \{x(1), x(2), ..., x(s)\}$, conditional of event m, $m = \{C, D\}$ and k_s is a time dependent alarm limit.

The event C (outbreak) is defined in accordance with which outbreaks we are interested in detecting. Sometime it is only of interest to detect whether there has been an outbreak at the current time point, i.e. whether the time of change is "now". This is expressed as $C = \{\tau = s\}$, where $\tau =$ time of outbreak, s=decision time. For this specification of C, the LR method above is reduced to only the likelihood ratio of the last observation, x(s). For our outbreak detection system we are instead interested in detecting all changes since the start of the surveillance, i.e. $C = \{\tau \leq s\}$, that is $C = \{C1, C2, ..., Cs\}$, where $Cj = \{\tau = j\}$. In this situation the LR method consists of the likelihood ratios of all observations, $\{x(1), x(2), ..., x(s)\}$,

$$LR(s) = \frac{f(x_s|C_1) \cdot P(C_1)}{f(x_s|D) \cdot P(D)} + \frac{f(x_s|C_2) \cdot P(C_2)}{f(x_s|D) \cdot P(D)} + ... + \frac{f(x_s|C_s) \cdot P(C_s)}{f(x_s|D) \cdot P(D)}.$$

The LR method is based on full knowledge of the process under surveillance, X, and of the change process (further discussed below). This means that the trend μ, conditional on C and D, is known. In a practical situation μ is not known, but can be estimated, for example from previous data. The estimation is often done under some parametric assumption, using observations from a window of the previous data. One possibility is to use the following parametric model for μ at time t

$$\mu(t): \begin{cases} \mu^D(t): \beta_0 + \beta_1 \cdot t, & t < \tau \\ \mu^{C\tau}(t): \beta_0 + \beta_1 \cdot (\tau-1) + \beta_2 \cdot (t-\tau+1), & t \geq \tau \end{cases},$$

where τ is the time of the outbreak.

The above specification of μ results in the following expression of the LR

$$LR(s) = \sum_{j=1}^{s} \frac{f(x_s|C_j) \cdot P(C_j)}{f(x_s|D) \cdot P(D)} = \sum_{j=1}^{s} \frac{f(x_s|\mu = \mu^{Cj}) \cdot P(C_j)}{f(x_s|\mu = \mu^{D}) \cdot P(D)}.$$

It is important to correctly specify the baseline (μ^D) and the outbreak (μ^C). In most cyclical processes it is unlikely that the length and shape of the cycles are constant over time. If μ is estimated from previous data, there is always the risk that the current cycle

is different from the previous ones, i.e. that μ be miss-specified when it is estimated from historic data. To avoid miss-specification, the full likelihood ratio is here approximated by the maximum likelihood ratio. The new in this approach is that no parametric restriction is placed on the cycles. The cycles are estimated using only monotonicity restrictions, see (9), (10). Only data from the current batch is used in the estimation, thus avoiding the risk of relying too much on the pattern of previous data. This approach, hereafter denoted the LRnp method (likelihood ratio, non-parametric), was suggested in (11) and evaluated in (12) and (13).

$$\text{LRnp(s)} = \sum_{j=1}^{s} \frac{f(x_s \mid \mu = \hat{\mu}^{Cj}) \cdot P(C_j)}{f(x_s \mid \mu = \hat{\mu}^{D}) \cdot P(D)}.$$

3.2 Intensity of the change-process

The optimal LR method includes the distribution for the change-point time, τ. Optimality has been proven for the case when τ has a geometric distribution. Much theoretic work on statistical surveillance is made under this assumption. The geometric distribution is also often assumed in HMM approaches where the transition probabilities are constant, see (1). The distribution of τ affects the weighting of the partial likelihood ratios as well as the alarm limit, as we can see when the maximum likelihood ratio method is expressed as

$$\text{LRnp} = \sum_{j=1}^{s} w_j \cdot \left(\frac{f(x_s \mid \mu = \hat{\mu}^{Cj})}{f(x_s \mid \mu = \hat{\mu}^{D})} \right) > k_s',$$

where $w_j = P(C_j) / P(C)$ and $k_s' = k \cdot P(D) / P(C)$.

What assumptions are reasonable to make about τ, the time of the outbreak? Is it possible to use previous data here? If we know that during previous influenza seasons the outbreak has started approximately 9 weeks after the start of the surveillance, does this information improve the performance? The pros and cons of using an empirical prior are evaluated in (14).

Many evaluations of the LR method is made for the situation when τ has a geometric distribution with parameter ν. When the limiting distribution of τ is used (when ν tends to zero), it is called the Shiryaev-Roberts approach, hereafter SR ((15), (16)). In the SR approach the weights for the partial likelihood ratios are equal and the alarm limit is constant. Using the SR approach can be expressed as having a non-informative prior for τ.

The effect of different assumptions regarding τ is evaluated. One of the evaluated systems is based on the distribution of the empirical change-point times, approximated by the negative binomial distribution. This is a very informative prior for τ. Another system is based on the Shiryaev-Roberts approach, thus a non-informative prior for τ.
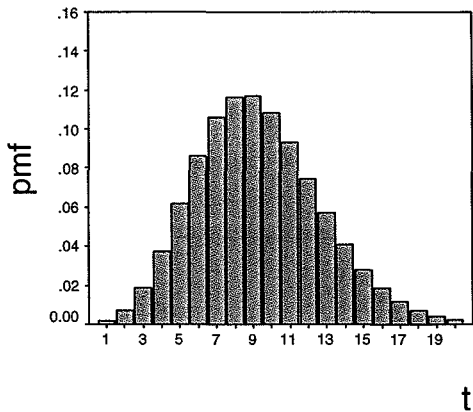
3

Fig 1: The density of the empirical distribution of the change point time is approximated by the negative binomial distribution

4. Results

4.1 False alarms

When an alarm system is evaluated it is important to know how often we can expect a false alarm. Here the alarm limit is set so that MRL0=12 (median run length to first false alarm). When setting the alarm limit, there is a trade-off between the false alarms and the delay of motivated alarms.
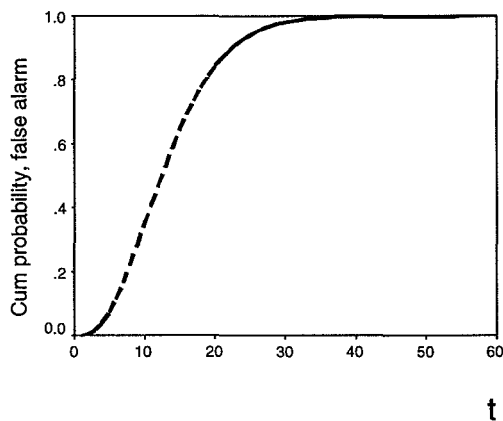

Fig 2: The false alarm distribution, when MRL0=12

As we can see, the probability of an alarm tends to one as the time tends to infinity. For methods where this probability does not tend to one, it is more difficult to give an alarm at late time points.

4.2 Motivated alarms

The probability of detection of motivated alarms within 2 time units is shown in Fig 3.
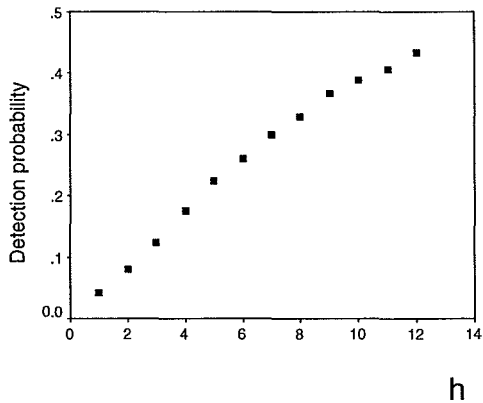
4

Fig 3: Probability of detection within 2 time units, as a function of when the outbreak starts in relation to start of surveillance

If the outbreak begins at the same time as the surveillance is started (h=1), then it is difficult to detect quickly (probability of detection is 0.04). This is because we have very little information on which to base our inference. If the outbreak instead begins after 12 time points, then the detection probability is considerable higher, 0.43.

4.3 Miss-specification of outbreak trend

One advantage with the non-parametric method LRnp is that the risk of miss-specification is reduced, since the estimation of the trend avoids using a parametric specification. Using parametric assumption regarding μ based on the previous pattern, there is always the risk that the current cycle is different from the previous ones, which results in a miss-specification of μ. In (17) it is demonstrated that a miss-specification of the trend after the turn is not so serious. But if both the trend before and after the turn, i.e. both the base line and the outbreak, are wrongly specified, it results in very bad properties for the alarm system.
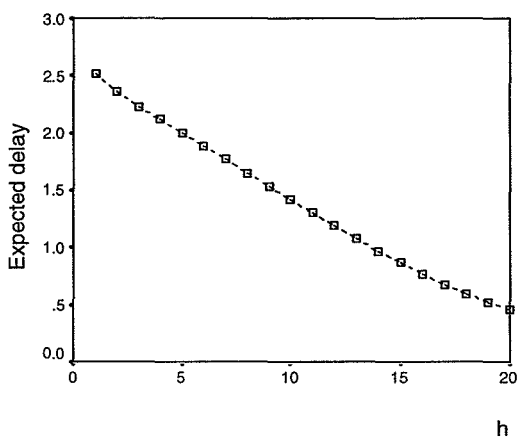

Fig 4: The expected delay of an alarm, as a function of when the outbreak starts in relation to start of surveillance. The situation is one where both the base line and the outbreak are miss-specified.

5

If the outbreak begins at the same time as the surveillance is started (h=1), then it takes a long time to detect it (the delay is 2.5). But if the outbreak begins after 20 time points, then the delay is small, 0.45.
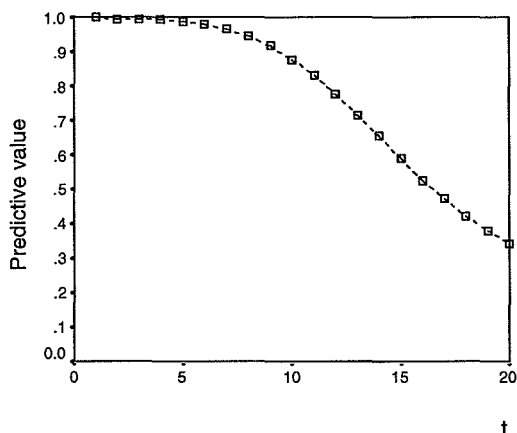

Fig 5: The predictive value of an alarm, as a function of when the outbreak starts in relation to start of surveillance. The predictive value is calculated under assumption of a constant intensity, 0.10. The situation is one where both the base line and the outbreak are miss-specified.

In Fig 5, we see that the short delay comes at a price: an alarm at time 20 is not very reliable, the predictive value is only 0.34. In order to calculate the predictive value, we must have some knowledge of the intensity of the change process and in Fig 5 the assumption of a constant intensity is used.

The above results, where both base line and outbreak are miss-specified, are compared to the non-parametric approach.
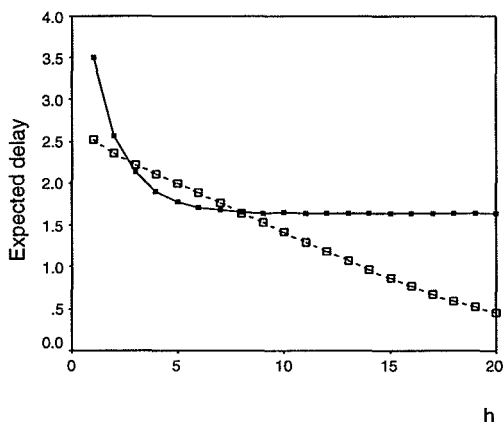

Fig 6: The expected delay, as a function of when the outbreak starts in relation to start of surveillance. The non-parametric approach (solid line, filled squares) is compared to a parametric approach with miss-specifications (dotted line, unfilled squares).
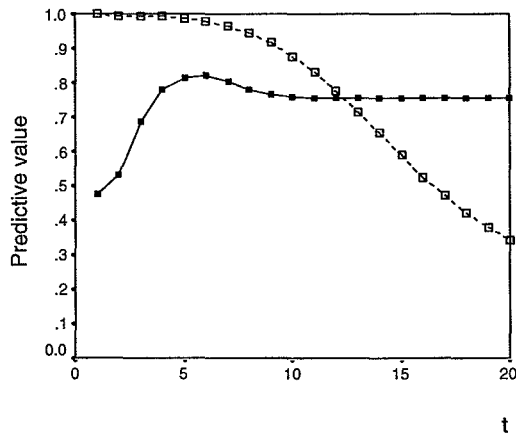
6

Fig 7: The predictive value, as a function of the time of the alarm. The predictive value is calculated under assumption of a constant intensity, 0.10. The non-parametric approach (solid line, filled squares) is compared to a parametric approach with miss-specifications (dotted line, unfilled squares).

4.4 Using prior information regarding the intensity

The weighting of the partial likelihood ratios means that more weight is given to those partial likelihood ratios that represent the most likely turning point. The non-informative prior gives equal weight to the all partial likelihood ratios, whereas the empirically based prior gives highest weight to that likelihood ratio which represents a turn at 9. The alarm limit is also affected by the choice of prior for τ. The non-informative prior results in a constant alarm limit, whereas the empirical prior has an extremely high alarm limit at early time points, resulting in few early alarms. In (14) the non-parametric method is evalauted for different priors regarding the intensity: the empirical prior is compared to a non-informative prior.
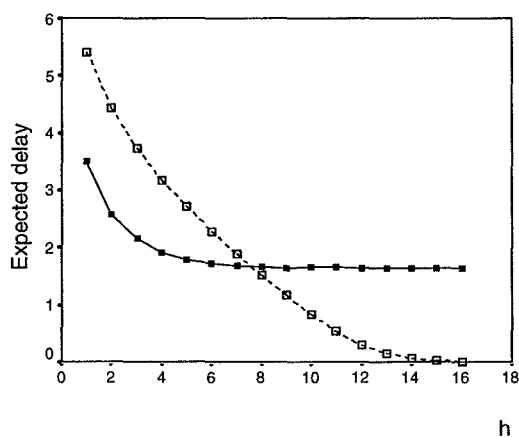

Fig 8: The expected delay, as a function of when the outbreak starts in relation to start of surveillance for the non-parametric method. A non-informativ prior for the intensity (solid line, filled squares) is compared to an empirical prior (dotted line, unfilled squares).

Using the empirical prior, the delay is very long, 5.4, for h=1, whereas the non-informative prior has a shorter delay, 3.5. If the outbreak does not begin at the same time as the surveillance, but instead at h=15, then the empirical prior has shorter delay, compared to the non-informative (0.02 and 1.64).
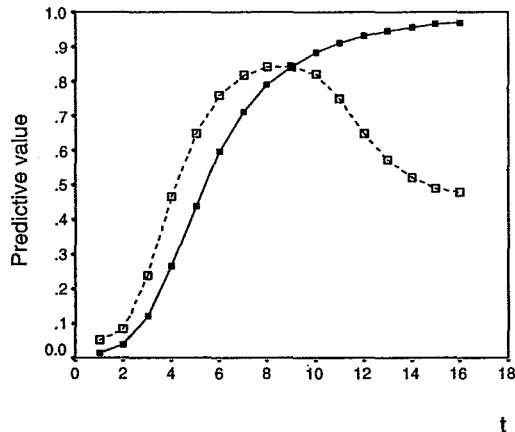

Fig 9: The predictive value, as a function of the time of the alarm, for the non-parametric method. The predictive value is calculated using the empirical prior for the change process. A non-informativ prior for the intensity (solid line, filled squares) is compared to an empirical prior (dotted line, unfilled squares).

But for the empirical prior the short delay at 15 is not of much use. The predictive value of an alarm at time 15 is only 0.49 for the empirical prior, whereas it is for the non-informative prior 0.96.

5. Conclusions

In surveillance it is important to use a method with known properties. Since timeliness is important, good properties in a surveillance system are measured by, for example, the probability to detect an outbreak within 2 time units from the start of the outbreak, or by the expected delay of an alarm. An optimal alarm system is based on the full likelihood ratio, which yields the shortest expected delay for a specified false alarm probability.

When a surveillance system is evaluated it is important that timeliness is considered. If the outbreak begins at the same time that the surveillance is started, then it is often more difficult to detect it, whereas if the outbreak begins later it is easier to detect. The predictive value of an alarm differs, depending on when the alarm is called. Often, an alarm called at the first time point carry less information than an alarm called later on. Thus, the extra dimension of time is needed in the evaluation.

With cyclical processes, such as influenza cycles, it is difficult to find a suitable parametric structure. Here an approximation of the full likelihood ratio is used, where the cycles are is estimated without any parametric assumptions regarding the shape. Only the assumption of unimodality is used, thus resulting in a very robust method. One advantage with the non-parametric approach is that miss-specification, especially of the base line, is avoided.

8

The question of how much prior information to use is often raised. If the intensity of the change process can be estimated from previous data, does this information improve the performance of the surveillance? Of course this depends on whether it is reasonable to believe that prior patterns in the change process will repeat themselves. The investigation shows that an empirical prior for the intensity works well when the actual outbreak time agrees with the prior. But when the outbreak occurs „earlier than expected", the alarms are seriously delayed. A non-informative prior, however, works well for all change point times.

REFERENCES

1. Hamilton, J. D. (1989) A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57, 357-384.
2. Royston, P. (1991) Identifying the fertile phase of the human menstrual cycle. *Statistics in Medicine*, 10, 221-240.
3. Kulldorff, M. (1997) A spatial scan statistic. *Communications in Statistics. Theory and Methods*, 26, 1481-1496.
4. Kulldorff, M. and Nagarwalla, N. (1995) Spatial disease clusters: Detection and inference. *Statistics in Medicine*, 14, 799-810.
5. Sonesson, C. and Bock, D. (2002) A review and discussion of prospective statistical surveillance in public health. *Journal of the Royal Statistical Society A*, 166, 5-21.
6. Frisén, M. (2003) Statistical measures for evaluation of methods for syndromic surveillance. *2003 National syndromic surveillance Conference*, New York, USA.
7. Frisén, M. and de Maré, J. (1991) Optimal Surveillance. *Biometrika*, 78, 271-80.
8. Le Strat, Y. and Carrat, F. (1999) Monitoring epidemiologic surveillance data using hidden Markov models. *Statistics in Medicine*, 18, 3463-3478.
9. Robertson, T., Wright, F. T. and Dykstra, R. L. (1988) *Order Restricted Statistical Inference*, John Wiley & Sons Ltd.
10. Frisén, M. (1986) Unimodal regression. *The Statistician*, 35, 479-485.
11. Frisén, M. (1994) Statistical Surveillance of Business Cycles. *Research report, Department of Statistics, Göteborg University, Sweden.*
12. Andersson, E. (2002) Monitoring cyclical processes. A non-parametric approach. *Journal of Applied Statistics*, 29, 973-990.
13. Andersson, E., Bock, D. and Frisén, M. (2002) Detection of turning points in business cycles. Conditionally accepted for publication in *Journal of Business Cycle Measurement and Analysis*.
14. Andersson, E. (2003) Surveillance of a cyclical process. The impact of intensity of turning points. *Research report, Department of Statistics, Göteborg University, Sweden.*
15. Shiryaev, A. N. (1963) On optimum methods in quickest detection problems. *Theory of Probability and its Applications*, 8, 22-46.
16. Roberts, S. W. (1966) A Comparison of some Control Chart Procedures. *Technometrics*, 8, 411-430.
17. Andersson, E., Bock, D. and Frisén, M. (2002) Statistical surveillance of cyclical processes with application to turns in business cycles. *Research report, Department of Statistics, Göteborg University, Sweden.*

Research Report

2002:7   Andersson, E.,   Some statistical aspects on methods for
         Bock, D. & Frisén, M.:   detection of turning points in business cycles.

2002:8   Andersson, E.,   Statistical surveillance of cyclical processes with
         Bock, D. & Frisén, M.:   application to turns in business cycles.

2002:9   Holgersson, T.:   Testing for non-normality in multivariate
                                   regression with nonspherical disturbances.

2003:1   Holgersson, T. &   Testing for multivariate heteroscedasticity.
         Shukur, G.:

2003:2   Holgersson, T.:   Testing for multivariate autocorrelation.

2003:3   Petzold, M. &   Detection of intrauterine growth restriction.
         Sonesson, C.:

2003:4   Bock, D.:   Similarities and differences between statistical
                                   surveillance and certain decision rules in
                                   finance.

2003:5   Holgersson, T. &   A comparison of conditioned versus
         Lindström, F.:   unconditioned forecasts of the VAR(1) process.

2003:6   Bock, D.:   Early warnings for turns in business cycles
                                   and finance.

2003:7   Lindström, F.:   On prediction accuracy of the first order vector
                                   auto regressive process.

2003:8   Petzold, M. &   Maximum Likelihood Ratio based small-sample
         Jonsson, R.:   tests for random coefficients in linear
                                   regression.

2003:9   Petzold, M.:   Preliminary testing in a class of simple non-
                                   linear mixed models to improve estimation
                                   accuracy.

2003:10   Frisén, M. &   Graphical evaluation of statistical surveillance.
         Gottlow, M.:

2003:11   Frisén, M.:   Statistical measures for evaluation of methods
                                   for syndromic surveillance.