STATISTISKA FORSKNINGSENHETEN, GÖTEBORGS UNIVERSITET
STATISTICAL RESEARCH UNIT, UNIVERSITY OF GOTHENBURG, SWEDEN

Skriftserie – Publications
32

Univariate and Multivariate Surveillance of Outbreaks

Linus Schiöler

UNIVERSITY OF GOTHENBURG

Thesis for the degree Doctor of Philosophy

Univariate and Multivariate Surveillance of Outbreaks

Linus Schiöler

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

Göteborg, 2010

UNIVERSITY OF GOTHENBURG

Univariate and Multivariate Surveillance of Outbreaks

Linus Schiöler

Statistical Research Unit
Department of Economics
University of Gothenburg
Sweden

ISBN 978-91-85169-54-2

Printed by Geson Hylte Tryck, Göteborg, Sweden 2010

Abstract

In many areas there is a need to monitor observations in order to detect changes in the underlying processes as quickly as possible. The theory of statistical surveillance provides the possibility of making optimal decisions about whether a change has occurred or not based on the data available at the time of the decision. Surveillance can be used in many different situations. It is important that the relevant characteristics of the change are identified and that the relevant optimality criterion is used. There is a need to further develop the theory of statistical surveillance.

One area where surveillance is of special interest is the detection of outbreaks of epidemic diseases. New strains of influenza virus like avian flu and swine flu have drawn much attention, but it is also important to detect the varying onset of the seasonal influenza. Outbreaks are characterized by a change from a constant incidence to an increasing one. A quick and reliable detection of epidemic outbreaks can be beneficial to society as it has the potential to prevent loss of lives and severe economic consequences. The detection of a change from a constant level to a monotonically increasing (or decreasing) regression is of interest also in other areas, for example in finance. This thesis considers outbreak detection in a wide meaning. It deals with topics of statistical surveillance in general and with applications to warning systems for influenza in particular.

When information on several variables is available it should be efficiently used in the surveillance system. The construction and evaluation of multivariate surveillance methods need to be developed, and one aim of the thesis is to contribute to this development.

In Paper I, a nonparametric univariate method for surveillance was applied to Swedish data on seasonal influenza and tularemia. An experiment to compare the statistical method to subjective judgment was performed. A user-friendly program implementing the method is presented.

As Swedish influenza data are collected from several different regions, a multivariate surveillance system could be superior to a univariate one. However, the evaluation of multivariate surveillance demands special care. Paper II deals with these problems. The suggested evaluation measures were subsequently used in Paper III and V.

In Paper III it was demonstrated that in some cases there exists a sufficient statistic that can be used to reduce a multivariate surveillance problem to a univariate one.

In Paper IV it was examined how the spreading pattern of influenza in Sweden could be characterized.

In Paper V, the information from the other papers was used to construct a method for multivariate outbreak detection. Motivated by the findings on the spreading pattern of influenza in Paper IV, the univariate outbreak detection method of Paper I was generalized to a multivariate method for outbreak detection by the results on multivariate techniques found in Paper II and Paper III.

Acknowledgements

First and foremost I would like to express my deepest gratitude to my supervisor Marianne Frisén for her support and guidance throughout the work with this thesis. She has generously devoted much time not only to giving me guidance into the world of statistical science and scientific thinking but also to giving encouragement and constructive suggestions.

I would also like to express my deepest thanks to my assistant supervisor Eva Andersson who has also given me invaluable support and guidance. Kjell Pettersson has given me many helpful comments during the work with the thesis as well as many insights into statistics in general.

I am grateful for the friendship and support from present and earlier members of the Statistical Research Unit and the department of Economics. Bitte Wallin has given expert linguistic advice.

I would like to thank my parents, my siblings and all my friends for support, encouragement, and many laughs. Finally I would like to thank Erika for brightening my day as well as for her encouragement and understanding during late work hours.

This thesis is based on five papers, referred to in the text by their respective Roman numerals:

I. Frisén, M., Andersson, E. and Schiöler L.
Robust outbreak surveillance of epidemics in Sweden.
Statistics in Medicine 2009, **28**, 476-493.

II. Frisén, M., Andersson, E. and Schiöler L.
Evaluation of Multivariate Surveillance.
To appear in *Journal of Applied Statistics.*

III. Frisén, M., Andersson, E. and Schiöler L.
Sufficient Reduction in Multivariate Surveillance.
To appear in *Communications in Statistics – Theory and Methods*.

IV. Schiöler, L.
Characterisation of influenza outbreaks in Sweden.
Submitted.

V. Schiöler, L. and Frisén M.
Multivariate outbreak detection.
Submitted.

1. Introduction

In many situations there is a need to repeatedly evaluate information as new observations of one or more variables become available. The methods of statistical surveillance can be used to decide as quickly and reliable as possible whether or not a change has occurred. One area of special importance, which will be explored in this thesis, is the surveillance of outbreaks of epidemic diseases. Early warnings of an outbreak may be vital for successful preventive action and may also aid in the planning of allocating resources to the primary care sector.

Statistical surveillance is termed differently in different areas. Many of these terms are broad and do not necessarily include statistical surveillance in the sequential sense discussed here. In industrial statistics the term SPC (statistical process control), or more specifically control charts, is often used. In medicine and economics the terms monitoring and early warning systems are common. In probability theory the term optimal stopping rules is frequent. Overviews on statistical surveillance can be found for example in Basseville and Nikiforov (1993), Frisén (2003), and Frisén (2009). Sonesson and Bock (2003) gave an overview on surveillance in public health.

As regards outbreaks, there are several definitions. One definition is a change in incidence from a constant level to an increasing one. This definition will be used in the thesis and is specified in the next section. Alternatively, an outbreak is commonly defined as an incidence higher than usual or as a spatial clustering but, as discussed in Paper I, these definitions were not found suitable for the applications in this thesis.

The need for multivariate surveillance occurs in many areas, such as industrial production, bioterrorism detection, spatial surveillance, and financial transaction strategies. This thesis will focus on spatial surveillance, which is a special case of multivariate surveillance as several data sources are used.

2. Models and specifications

This thesis deals with the surveillance of discrete time points, as is most common in surveillance of epidemics. The process under surveillance will be denoted $Y(t)$ for the univariate case and $\mathbf{Y}(t)$ for the multivariate case, where $\mathbf{Y}(t)$ denotes the p-variate vector observed at time t, i.e. $\mathbf{Y}(t) = \{Y_1(t), Y_2(t), ..., Y_p(t)\}$. For a univariate process, as studied in Paper I, an outbreak will be defined as a change in expected values, $m(t)$, from a constant level to a monotonically increasing one

$$\begin{aligned} &\mu(1) = \mu(2) = ... = \mu(s) && \text{for } \tau > s \\ &\mu(1) = ... = \mu(\tau - 1) < \mu(\tau) \le ... \le \mu(s) && \text{for } \tau \le \text{s} \end{aligned} \qquad (1)$$

where τ is the unknown time of the onset of the outbreak and s is the time of the decision.

In the multivariate case, as studied in Paper V, each of the p components of the process may change at different times. Hence, the definition of each outbreak is analog to the univariate case; for each of the p processes Y_i there is a change in expected value at the unknown time τ_i. The processes are assumed to be independent given the values of τ_i.

In statistical surveillance the aim is to repeatedly evaluate information and at each new time point s discriminate between different states of the process, for example the two states of an outbreak as defined in (1). In multivariate surveillance there are several change points. The

aim is to detect any change in the system. Thus, the change in the process which changes first is of special interest.

Surveillance systems based on likelihood ratios are optimal under many criteria. In order to use the likelihood ratio the distribution of the process under surveillance must be known. In Andersson *et al.* (2008) it was found that a simple stochastic model without serial correlation but with time dependent expectation is suitable for the surveillance of influenza in Sweden. The results also showed that the Gaussian distribution is suitable for observations near the peak while the Poisson distribution is suitable at the onset. Most of the theory in this thesis is derived for the one parameter exponential family and for the two parameter family with known dispersion parameter. Since the emphasis of the thesis is on the onset phase, the Poisson distribution will be used for specific results.

In Paper IV, the spreading pattern of influenza in Sweden was shown to be characterized by a time lag in the onset in different regions. In Paper V, this time lag in the multivariate model was used to construct a system for monitoring influenza in Sweden.

3. Methods for surveillance

The earliest methods for surveillance were developed by Shewhart (1931) for the control of industrial processes. The full likelihood ratio method was derived in Shiryaev (1963). Other popular surveillance methods include the CUSUM method (Page 1954), the EWMA method (Roberts 1959) and the Shiryaev-Roberts method (Shiryaev 1963, Roberts 1966). Most methods can be expressed by different combinations of the partial likelihood ratios and correspond to different optimality criteria.

In Paper I and V, the change situation was partially known and partially unknown. The distribution function of the process was known to belong to the exponential family. The change was one from a constant expected value to an increasing expected value, but there was no parametric assumption on the shape of the increase. A generalized likelihood ratio approach and ordered restricted inference were used to handle the unknown parameters for both the univariate situation in Paper I and the multivariate situation in Paper V. For the multivariate situation this was combined with a sufficient reduction derived by a modification of the sufficiency results of Paper III.

4. Evaluation in statistical surveillance

In surveillance there is always a trade-off between making a fast decision and avoiding making an incorrect one. There are several metrics used to measure both timeliness and correctness. A further discussion can be found for example in Frisén (1992) and Knoth (2006).

The intricate issue of evaluation in multivariate surveillance is treated in Paper II. It is important to clearly specify the aim, since there are possibly multiple change points.

4.1. False Alarms

A commonly used measure of false alarms is the in-control average run length, ARL^0, $E[t_A|\tau=\infty]$ where t_A is the time of alarm. A similar measure, which is more convenient to calculate and used here, is the median run length, MRL^0. This measure is also less sensitive to skewed distributions of the alarm times. Both measures can be used also in a multivariate situation by specifying that no change occurs in any of the processes.

4.2. Delay

One measure of the detection ability is $E[t_A| \tau=1]$, the average run length given that the change occurs immediately. This is often named ARL[1] or zero-state ARL and is widely used in univariate surveillance. It is commonly used as an evaluation measure also in the multivariate case. However, it is seldom explicitly defined. The definition implicit in most publications is $E[t_A| \tau_1= \tau_{2= \dots} \tau_p =1]$. Thus, it is implied that all processes change at the same time. For this case a sufficient reduction to a univariate problem exists, as demonstrated in Paper III. Zero-state ARL is thus questionable as a formal measure for comparing methods for handling genuinely multivariate problems. Instead, a measure which allows different change points was used, namely a multivariate version of the conditional expected delay.

The conditional expected delay, $CED(\tau) = E[t_A - \tau \mid \tau \le t_A]$, which was used for the univariate case of Paper I, can be generalized to $CED(\tau_1, \tau_{2\dots} \tau_p) = E[t_A - \tau_{\min} \mid \tau_{\min} \le t_A]$ for multivariate surveillance, as in Paper II, III, and V.

4.3. Predictive value

If a method calls an alarm it is important to know whether this alarm is a strong or weak indication of a change. In epidemiology, the predictive value is a well-established measure. In surveillance, however, there is need for a variant that also incorporates time. The difference between surveillance and situations involving only one decision is that in surveillance there can be an alarm at any time point, and therefore a measure of the predictive value at each possible time point is needed. In order to judge to what degree an alarm at time t_A can be trusted, it is necessary to consider the balance between the risk of false alarms, the detection ability, and the probability of a change. If there is one change point τ and this is regarded as a random variable, the predictive value is calculated as the probability of an outbreak conditional on an alarm, as suggested by Frisén (1992). In Paper II, this approach was extended to the multivariate case, and it was subsequently used in Paper V.

5. Conclusions

In recent years, there have been several events that highlight the importance of outbreak detection. The outbreaks of new kinds of influenza (SARS, avian flu and H1N1) are such recent examples.

The choice of outbreak detection method and evaluation procedure depends on which definition of outbreak is used. Therefore, it is important to state the aim explicitly. Different methods may be optimal under different conditions, which means that the methods can often be seen as complements to each other. The methods developed and employed in this thesis treat outbreaks defined as a monotonic increase following the constant level before the onset of the outbreak. Such outbreaks are of interest in connection with several diseases and syndromes. Often, the information about the baseline is limited. Errors in the estimation of the baseline can have serious effect, as demonstrated for example by Frisén and Andersson (2009). Also, there may be seasonal effects with the same periodicity as the disease as well as large variation between years, thus making it hard to estimate the expected baseline incidence. Thus a semi parametric method was chosen, as it does not require information about the baseline.

When data from different sources are available, multivariate surveillance should be applied. One such example is the detection of influenza outbreaks on the basis of data from different regions. The two simplest approaches of multivariate surveillance are the reduction to a suitable univariate statistic and the surveillance of the separate processes in parallel. In

multivariate surveillance the properties of a method depend heavily on the relation between the times of onset in the different processes. In this thesis, Swedish data from metropolitan regions and local regions (where the outbreaks occurred at different times) were combined in a system for influenza outbreak detection. The information on the delay in the outbreak between the two types of regions was the base for a new approach of multivariate surveillance, and the results of this method were compared to conventional approaches. The relation between the different processes is important in multivariate surveillance, as demonstrated in Paper II. A method that is optimal for simultaneous changes is not efficient when there is a time lag. The exact relation between the onsets at different locations is seldom exactly known. However, some information may be available, as demonstrated in Paper IV, where it was found that the influenza outbreak in Sweden generally started a week earlier in the major cities than in the rest of the country. In the application to Swedish influenza data in Paper V, it was demonstrated that the performance of the surveillance method was improved by utilizing this knowledge. A joint generalized likelihood ratio method, based on maximum likelihood under multivariate monotonicity restrictions and a sufficient combination of data, was suggested and utilized. In simulation studies and when applied to Swedish influenza data, the multivariate method performed better than other methods described in the paper. It was also demonstrated that if the true time lag is only approximately known, the results can be improved by using this information in the surveillance procedure.

When evaluating methods for on-line monitoring it is important to use measures that incorporate the issue of time, i.e. the fact that there are repeated decisions, not just one decision as in hypothesis testing. This is even more important in the multivariate case where there are several possible change points. In Paper II, evaluation measures which are better suited for multivariate on-line surveillance than the conventional ones are suggested.

The method of Paper I has recently been implemented in the open source JAVA package CASE, which is currently in use at the Swedish Institute for Infectious Disease Control, see Cakici *et al.* (2010) for further details. It has also been implemented in the R package Surveillance, which is described in Höhle (2010).

6. Summaries of the papers

6.1. Paper I: Robust outbreak surveillance of epidemics in Sweden

A semi parametric method for outbreak detection was applied to Swedish data on tularemia and influenza. The method was constructed to detect a change from a constant level to a monotonically increasing incidence. The properties of the method were evaluated by the applications and by simulations. The suggested method was compared with subjective judgments as well as with other algorithms. The conclusion was that the method works well compared to both subjective judgment and other algorithms. The method was implemented in a user-friendly computer program, which is described.

6.2. Paper II: Evaluation of Multivariate Surveillance

There are many measures of the performance of univariate surveillance methods, for example expected delay. Special care is needed when using these measures in multivariate surveillance. In Paper II, some new measures are suggested, and their properties were investigated by applications to various situations. It was demonstrated that the delay is dependent on the time differences between the change points, and hence that measures such as zero-state ARL and steady state ARL should be used with care.

6.3. Paper III: Sufficient Reduction in Multivariate Surveillance

The relation between change points in a multivariate process under surveillance is important but seldom considered. The sufficiency principle was used both to clarify the structure of some problems and to find efficient methods as well as to determine appropriate evaluation metrics. Processes where the changes occurred simultaneously or with known time lags were studied. A general version of a theorem for the sufficient reduction of processes that change with known time lags is given. A simulation study illustrated the benefits of the methods based on the sufficient statistics.

6.4. Paper IV: Characterisation of influenza outbreaks in Sweden

In Paper IV, spatial aspects of the seasonal Swedish influenza outbreak were investigated. In Paper I, a semi parametric surveillance method was applied to data on Sweden as a whole. Here, however, the aim was to study the data on individual regions to determine whether there was a spatial pattern that could make it feasible to employ a multivariate surveillance method. Quality problems were found for all types of available data. The results of the analyses showed that the outbreak usually starts about a week earlier in the major cities with large airports compared to the rest of the country. It was suggested that nonparametric methods be used for inference and surveillance. A bivariate parametric model was suggested for simulation purposes.

6.5. Paper V: Multivariate outbreak detection

In Paper V, the sufficient reduction derived in Paper III was extended to the case of an increasing level after the change point. The robust semi parametric surveillance method in Paper I was adopted to use the sufficient reduction. This new method was evaluated with respect to robustness and efficiency in a simulation study. The result showed that if an approximately known time lag exists, the method should be considered. The method was also applied to spatial data for detection of influenza outbreaks in Sweden, where the aggregation procedure suggested in Paper IV was used. In this application the new method performed better than the univariate one.

References

Andersson, E., Bock, D. and Frisén, M. (2008) Modeling influenza incidence for the purpose of on-line monitoring. *Stat. Methods Med. Res.,* 421-438.

Basseville, M. and Nikiforov, I. (1993) *Detection of Abrupt changes - Theory and Application*, Englewood Cliffs: Prentice Hall.

Cakici, B., Hebing, K., Grünewald, M., Saretok, P. and Hulth, A. (2010) CASE –a framework for computer supported outbreak detection. *BMC Medical Informatics and Decision Making*.

Frisén, M. (1992) Evaluations of methods for statistical surveillance. *Stat. Med.,* 1489-1502.

Frisén, M. (2003) Statistical surveillance. Optimality and methods. *Int. Stat. Rev.,* 403-434.

Frisén, M. (2009) Optimal sequential surveillance for finance, public health and other areas. Editor's special invited paper. *Sequential Analysis,* 310-337, discussion 338-393.

Frisén, M. and Andersson, E. (2009) Semiparametric surveillance of monotonic changes. *Sequential Analysis,* 434-454.

Höhle, M. (2010) Aberration Detection in R Illustrated by Danish Mortality Monitoring In *Biosurveillance* (eds T. Kass-Hout and X. Zhang): CRC Press.

Knoth, S. (2006) The art of evaluating monitoring schemes - how to measure the performance of control charts? In *Frontiers of Statistical Quality Control* (eds H.-J. Lenz and P.-T. Wilrich), pp. 74-99. Warsaw: Physica Verlag.

Page, E. S. (1954) Continuous inspection schemes. *Biometrika,* 100-114.

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

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

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

Shiryaev, A. N. (1963) On optimum methods in quickest detection problems. *Theory of Probability and its Applications.,* 22-46.

Sonesson, C. and Bock, D. (2003) A review and discussion of prospective statistical surveillance in public health. *J. R. Statist. Soc. A,* 5-21.

Paper I

STATISTICS IN MEDICINE
Statist. Med. 2009; **28**:476–493
Published online 14 November 2008 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/sim.3483

Robust outbreak surveillance of epidemics in Sweden

M. Frisén[1,*,†], E. Andersson[1,2] and L. Schiöler[1]

[1]*Statistical Research Unit, Department of Economics, University of Gothenburg, SE 40530, Gothenburg, Sweden*
[2]*Department of Occupational and Environmental Medicine, Sahlgrenska University Hospital, SE 40530, Gothenburg, Sweden*

SUMMARY

Outbreak detection is of interest in connection with several diseases and syndromes. The aim is to detect the progressive increase in the incidence as soon as possible after the onset of the outbreak. A semiparametric method is applied to Swedish data on tularaemia and influenza. The method is constructed to detect a change from a constant level to a monotonically increasing incidence. If seasonal effects are present, the residuals from a model incorporating these can be used. The properties of the method are evaluated by application to Swedish data on tularaemia and influenza and by simulations. The suggested method is compared with subjective judgments as well as with other algorithms. The conclusion is that the method works well. A user-friendly computer program is described. Copyright © 2008 John Wiley & Sons, Ltd.

KEY WORDS: exponential family; influenza; monitoring; ordered regression; subjective judgment; tularaemia

1. INTRODUCTION

Epidemic diseases cause much suffering to individuals and also have negative consequences for the society as a whole. Recently, there has been much interest in early warnings for outbreaks. Preparedness for pandemics is important, as stressed in [1, 2]. New or recently introduced diseases, such as SARS and avian flu, are of great concern. Acts of bioterrorism might cause (or look like) epidemic diseases.

International, national, and local institutes (for example, the Swedish Institute for Infectious Disease Control, SMI) collect data on several diseases and give timely information to the local

*Correspondence to: M. Frisén, Statistical Research Unit, Department of Economics, University of Gothenburg, SE 40530, Gothenburg, Sweden.
†E-mail: marianne.frisen@statistics.gu.se

Contract/grant sponsor: Swedish Emergency Management Agency; contract/grant number: 0314/206

Received 1 April 2008
Accepted 25 September 2008

society as well as to international organisations and networks (for example, the European Influenza Surveillance Scheme, EISS). A description of the present Swedish system for collecting data on influenza and influenza-like illness can be found for example in [3–5]. Much work is being done all over the world to improve the collection of data, which will increase the capacity for early detection of epidemics.

Predictions of future incidences at an epidemic outbreak play an important role for the planning carried out by health authorities. Predictions of the height and time of the influenza peak based on early observations were described in [4]. For several diseases and syndromes, however, it is important to detect the outbreak of an epidemic and to detect it as soon as possible after the onset. Outbreak detection is an inferentially different problem compared with prediction. In this paper, we define outbreak detection as on-line detection: each week when we make a new observation on the incidence series, we decide whether or not we have enough information to state that the influenza epidemic has started, i.e. that the increase in incidence is larger than what could be expected during the non-epidemic season.

There are three major approaches to outbreak detection: (i) the detection of an increasing incidence, (ii) the detection of an incidence that is higher than expected, based on the information available up to that point and (iii) the detection of spatial clustering of cases, which results in an uneven incidence. In the latter approach, the outbreak is characterized by a large number of individuals spatially close to each other getting ill approximately at the same time, while there is no increase in the surrounding areas. If spatial information is available, for example, the proportion of sick individuals in each municipality, the outbreak can be detected by a cluster detection method. One example is the Satscan method by Kulldorff [6], which was evaluated in a CUSUM framework in [7]. Many variants have been suggested. For example, Rosychuk *et al.* [8] included the case where the population size may vary by using a compound Poisson distribution. If, however, the outbreak is characterized by a general increase in incidence in large areas, cluster detection may not work better than the detection of a general increase. The spatial pattern at an outbreak is not always a circular spread, and other spatial approaches can thus be of interest, as discussed in [9]. No simple spread pattern could be found, such as a spread between neighbouring areas. The spatial component was more complicated and the start of the influenza epidemic was nearly simultaneous in the four major cities, which are far apart geographically. In this paper, we suggest a monitoring method that is based on aggregated data for the whole of Sweden. Information about data for different geographical areas could be of value and spatial surveillance is then important, [10]. Our suggested surveillance method may be used together with spatial information in such a context. However, we will not further discuss spatial issues here. In this paper, we use the first approach, i.e. the detection of an increasing incidence.

The most commonly used approach to detect an increased incidence described in scientific journals is based on a parametric model for the non-epidemic periods. A signal is given as soon as one observation exceeds a threshold, usually a 95 per cent prediction interval (see e.g. [11]). This is a variant of the Shewhart surveillance method, which is described in Section 2.2. This is not always an efficient method, as only the last observation is utilized. In [12, 13] it is suggested that there should be an alarm as soon as there are two consecutive observations beyond the limit. A more fruitful approach could be the use of a surveillance method, which gives optimal weights to the observations.

In [14] a hidden Markov model is suggested, which allows switching between states with different statistical properties: the non-epidemic state, with a low-level on the incidence rates, and the epidemic state, characterized by an increased incidence. The cyclical regression suggested

DOI: 10.1002/sim

in [12] was used to model the seasonal effect. It was found that the seasonality had a period of 52 weeks.

In [3] it is concluded that parametric methods are not suitable when the parameters describing the incidence curve vary much from year to year, as is the case in Sweden. Thus, we suggest that a semiparametric approach by Frisén and Andersson [15] should be used. The suggested method aims at detecting a change from a constant level to an increasing function, i.e. a change in monotonicity. It is a likelihood-based surveillance method. The suggested method is applied to Swedish data on tularaemia and common influenza.

In [16], several monitoring methods are described and compared: SPOTv2 (see [17]), the England–Wales method (see [18–21]), and two versions of the CUSUM method originally suggested by Page (see [22]). The suggested method is compared with these methods by conventional measures like sensitivity and positive predictive value. The method is also evaluated for detection of influenza outbreaks by other measures that take into account timeliness and the fact that there are repeated decisions.

In Section 2, a short background on the statistical surveillance of outbreaks is given. First, in Section 2.1, we describe models and specifications. In Section 2.2, we describe general likelihood-based methods for surveillance and the suggested semiparametric method, which was derived from an optimal likelihood ratio method. In Section 2.3 we give arguments for a semiparametric approach. In Section 2.4, a newly developed user-friendly computer program for outbreak detection is presented. Evaluation measures are given in Section 2.5. In Section 3, the suggested method is applied to Swedish data on tularaemia, and the result is compared with those of several other methods by conventional measures. In Section 4, the suggested method is applied to Swedish influenza data. A Monte Carlo simulation and a comparison with subjective judgments are used to evaluate the method. In Section 5 conclusions are given.

2. STATISTICAL SURVEILLANCE OF OUTBREAKS

2.1. Specifications of the outbreak problem

We monitor the process X and we observe $x(t)$ in discrete time. X could be a measure of the severity of a disease, for example an incidence. In general, frequent data contain more information than sparse. However, for many diseases, weekly data are what is practically available. Also, with weekly data the problem of influence of weekday is avoided. For the suggested method, it is not necessary that the time intervals between the observations are the same. The only requirement is that there is a natural ordering between the times of the observations. The values $t=1,2,\ldots,$ reflect the ordering of the times, but are not necessarily the time values that are natural in the application.

Let τ be the time of the start of the outbreak. The expected value of $X(t)$ conditioned on $\tau=i$ is denoted by $\mu^i(t)$. The superscript is suppressed when obvious. The decision time is denoted by s. At each decision time s we want to discriminate between the two events C and D, where $C=\{\tau\leqslant s\}$ 'the outbreak has occurred', and $D=\{\tau>s\}$ 'the outbreak has not occurred'. This discrimination is made by using the available observations, $x_s=\{x(1),\ldots,x(s)\}$. In the current outbreak situation we specify C and D in terms of a change in the expected value, so that for $s<\tau$ we have $\mu(t)=\mu^D(t)$ and for $s\geqslant\tau$, $\tau=j$ we have $\mu(t)=\mu^{C_j}(t)$. At an outbreak at the unknown time τ, the expected value μ changes from a constant (baseline) level to an increasing curve.

DOI: 10.1002/sim

If we know the parametric model for μ, this should be used to make the surveillance more efficient. The parametric model could be a change from a known baseline level (μ_0) to a (known) exponentially increasing function, so that

$$\begin{aligned} &\text{for } t<\tau \quad \mu(t)=\mu_0 \\ &\text{for } \tau=j, t\geqslant j \quad \mu(t)=\exp(\beta_0+\beta_1(t-j+1)) \end{aligned} \tag{1}$$

where μ_0, β_0, and β_1 are known constants.

However, if no certain information about the parametric shape of the curves is available, the general characteristics of an outbreak can be used. We study outbreaks which, at the decision time s (the latest observation point), are characterized by

$$\begin{aligned} &\text{for } s<\tau \quad \mu(1)=\mu(2)=\cdots=\mu(s) \\ &\text{for } s\geqslant\tau=j \quad \mu(1)=\mu(2)=\cdots=\mu(j-1)<\mu(j)\leqslant\cdots\leqslant\mu(s) \end{aligned} \tag{2}$$

For $s<\tau$ no outbreak has started yet, since the time of the change, τ, is in the future (this is state D). For $\tau=j\leqslant s$, an outbreak has started at time j (this is state C_j).

For the stochastic variable X, different distributions can be relevant for different diseases. As regards influenza, the conclusion in [3] was that independent Poisson distributions were suitable for the outbreak observations. The use of such distributions in outbreak detection is further discussed in Section 4.1.

2.2. General likelihood-based surveillance methods

Some characteristics separate a surveillance situation from a hypothesis testing situation. In hypothesis testing, we use the sample data to perform one test of whether a fixed null hypothesis can be rejected or not. In on-line surveillance, an alarm statistic is calculated at each new time point, for example, each week when new data become available. We make repeated decisions to determine whether the process is in state D or if it has changed to state C, see (2). This differs from hypothesis testing since the specification of both D and C changes as time progresses. In hypothesis testing, we evaluate the performance of a test by the power for a fixed size. In on-line surveillance, we often want to use measures that reflect the timeliness of the alarms (this is further discussed in Section 2.5).

Here, the process under surveillance, X, is the incidence of a disease or syndrome during each week (or some other period). One early method for the on-line surveillance is the Shewhart method, [23], for which an alarm is called the first time s that

$$(x(s)-\mu^D(s))>k_s \tag{3}$$

where $\mu^D(s)=E[X(s)|D]$ and k_s is the alarm limit (often $3*\sigma$). $\mu^D(s)$ is sometimes called the baseline, but it need not be constant over time. For the Shewhart method, μ^D is assumed to be known. The Shewhart method is optimal for the situation when we want to discriminate between $C=\{\tau=s\}$ and $D=\{\tau>s\}$. Here, however, we are interested in detecting whether the onset has occurred at any time during our surveillance, i.e. we want to discriminate between $C=\{\tau\leqslant s\}$ and $D=\{\tau>s\}$. Thus, the Shewhart method is not optimal here.

Shiryaev [24] showed that for discriminating between $C=\{\tau\leqslant s\}$ and $D=\{\tau>s\}$, the full likelihood ratio between C and D is optimal in the sense that the method gives a minimal expected

DOI: 10.1002/sim

delay for a fixed false alarm probability. The full likelihood ratio method is expressed as

$$\frac{f(x_s|C)}{f(x_s|D)} = \sum_{j=1}^{s} w_{js} \frac{f(x_s|\tau=j)}{f(x_s|\tau>s)} \tag{4}$$

where f is the likelihood function, $x_s = \{x(1), \ldots, x(s)\}$ and weights $w_{js} = P(\tau=j)/P(\tau \leqslant s)$.

If a parametric approach had been used, then μ would be specified, conditional on C and D. The alarm rule is

$$\mathrm{LR}(s) = \sum_{j=1}^{s} w_{js} \frac{f(x_s|\mu=\mu^{Cj})}{f(x_s|\mu=\mu^{D})} > k_s$$

where k_s is proportional to $P(\tau<s)/P(\tau \geqslant s)$. In the Shiryaev–Roberts approach [24, 25] the weights, w_{js}, are constant and so is the alarm limit.

2.3. The semiparametric approach

Many surveillance methods are based on the deviation between the observation and a baseline value (for example, between the number of influenza cases and μ_0). If we want to detect a deviation from a parametrically specified state D, this state must be well known. Also important, but not to the same extent, is that the change (state C) should be specified. As regards influenza, however, the characteristics of the outbreak vary from one year to the next. Thus, it is difficult to find a parametric model that describes the baseline and the outbreak of every year correctly (here, to find the correct parameter values for μ_0, β_0, and β_1 in (1)). It could be argued that the parameters could be sequentially updated, but this will not solve the fundamental problem that the current influenza season may differ substantially from the 'average' one, which is captured by the updated parameter estimates. Thus, the main objection to using a parametric surveillance method is that it is difficult to estimate the baseline model with any certainty. As was shown in [15, 26], a misspecified in-control model will result in an alarm system with poor performance.

A surveillance method based on the maximum likelihood ratio was suggested in [27] for the situation where we want to detect a turning point (peak or trough). The method was based on non-parametric estimation under monotonicity restrictions [28] as well as the use of the maximum likelihood ratio in the Shiryaev–Roberts approach. The maximum likelihood-based surveillance method was used in [29] in connection with peak detection regarding influenza.

Another maximum likelihood-based surveillance method was suggested in [15]. This method is based on the monotonicity restrictions in (2) and has been derived for both the normal and Poisson distributions, see [15]. For the Poisson distribution this method, OutbreakP, calls an outbreak alarm when

$$\mathrm{OutbreakP}(s) = \prod_{t=1}^{s} \left(\frac{\hat{\mu}^{C1}(t)}{\hat{\mu}^{D}(t)} \right)^{x(t)} > k \tag{5}$$

In (5), k is a constant alarm limit, and $\hat{\mu}^{C1}(t)$ and $\hat{\mu}^{D}(t)$ are the maximum likelihood estimates under the restrictions in (2). These maximum likelihood estimates are derived in [30], where it is shown that $\hat{\mu}^{D}(t)$ is the average of all the observations and that $\hat{\mu}^{C1}(t)$ is the isotonic regression estimator, which can be calculated by the pool adjacent violator algorithm (PAVA) described by e.g. [31]. The computation is illustrated by an example below. It is not possible to base the alarm

DOI: 10.1002/sim

Table I. An example of the computation at the first decision time.

t	$x(t)$	$\hat{\mu}^D(t)$	$\hat{\mu}^{C1}(t)$	$\left(\frac{\hat{\mu}^{C1}(t)}{\hat{\mu}^D(t)}\right)^{x(t)}$
1	11	10	10	1
2	9	10	10	1

Table II. An example of the computation at the second decision time.

t	$x(t)$	$\hat{\mu}^D(t)$	$\hat{\mu}^{C1}(t)$	$\left(\frac{\hat{\mu}^{C1}(t)}{\hat{\mu}^D(t)}\right)^{x(t)}$
1	11	20	10	0.5^{11}
2	9	20	10	0.5^{9}
3	40	20	40	2^{40}

statistic on a single observation, since the maximum likelihood method uses the ordering of the data. Thus, the first decision is taken when we have two observations. To stress that this is at time 2 we call this decision time $s=2$ (Table I).

The maximum likelihood estimate of the constant level, $\hat{\mu}^D(t)$, is calculated as the average of the two observations. The two observations violate the monotonicity restriction and therefore, by the PAVA, $\hat{\mu}^{C1}(t)$ is calculated as the average of these. The alarm statistic at the first decision time has the value

$$\text{OutbreakP}(2)=\prod_{t=1}^{2}\left(\frac{\hat{\mu}^{C1}(t)}{\hat{\mu}^D(t)}\right)^{x(t)}=1\cdot 1=1$$

At the next decision time, we have the details given in Table II.

Again, $\hat{\mu}^D(t)$ is calculated as the average. Here $x(1)$ and $x(2)$ violate the monotonicity restriction and are therefore pooled, whereas $x(3)$ does not and is kept. Thus, we have

$$\text{OutbreakP}(3)=\prod_{t=1}^{3}\left(\frac{\hat{\mu}^{C1}(t)}{\hat{\mu}^D(t)}\right)^{x(t)}=0.5^{11}\cdot 0.5^{9}\cdot 2^{40}=1\,048\,576$$

For long series of data, the calculations are best made with a computer program as described in the next section. The choice of the value of k (the alarm limit) depends on which properties are desired for the specific application. Measures for evaluations are discussed in Section 2.5.

2.4. User-friendly computer programs

In order to make the new methodology easily accessible, user-friendly computer programs have been developed. The programs, which are freely available from the authors, were developed using 'Visual Basic for Applications' in Microsoft Excel. The program basically looks like an ordinary Excel workbook, but has an additional user interface and algorithms to calculate the alarm statistic of OutbreakP. The user interface consists of a menu for running the program, dialogs for entering

DOI: 10.1002/sim

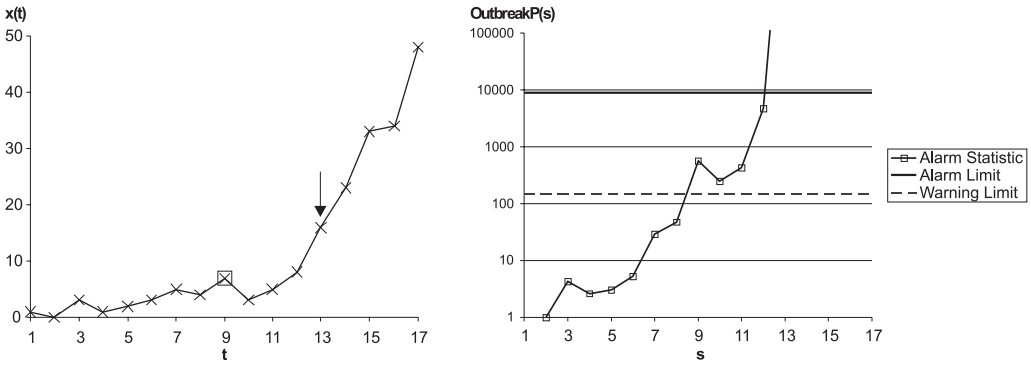

Figure 1. Left: number of laboratory confirmed cases of influenza in Sweden for the 07–08 season. The first observation is for week 40 2007. The warning and alarm signals of the method are marked by the square and arrow, respectively. Right: the output from the computer program corresponding to the data to the left.

limits, and a help screen. At each time point, the latest observation is added and the program produces the value of the alarm statistic as well as a warning or an alarm signal if there is enough evidence of an outbreak. All this output is displayed within the workbook. Graphs of the data and the regression under the outbreak model (2) are given. Graphs are also produced of the alarm statistic (as in Figure 1) in order to make it easy to compare with a warning or alarm limit after each new entry of weekly data.

2.5. *Evaluation measures*

Some of the evaluation measures commonly used in epidemiological studies are sensitivity, specificity, and positive predictive value. These measures were originally developed for the retrospective analysis of a data set where the outbreak has either occurred (state C) or not (state D). In such a retrospective setting, there is only one possibility of making an alarm (A); we make only one decision (as opposed to on-line monitoring with repeated decisions). There is also only one possible time point at which to not make an alarm ($\bar{A}$). The measures are defined as

$$\text{sensitivity} = P(A|C) \tag{6}$$

$$\text{specificity} = P(\bar{A}|D) \tag{7}$$

$$\text{positive predictive value} = \text{PPV} = P(C|A) \tag{8}$$

In Section 3, we will compare the OutbreakP method with the results of a study where evaluations were conducted by using the measures (6)–(8). In our comparison, we will use the same measures. A drawback of these measures in on-line surveillance is that they do not incorporate the timeliness of alarms. They were developed for the situation where only one decision is to be made, not for the repeated decisions required in surveillance. For example, the specificity depends on how long the surveillance has been in operation. It will tend to zero as the time progresses.

DOI: 10.1002/sim

For outbreak detection, we usually want evaluation measures that reflect timeliness. Thus, the time of the alarm is very important. It is defined as

$$t_A = \min[t : h(x_t) > g(t)]$$

where $h(x_t)$ is the alarm statistic and $g(t)$ is the possible time-dependent alarm limit. Measures developed for the on-line surveillance situation are the average run length to the first false alarm (ARL^0), the conditional expected delay of an alarm when the onset of the outbreak occurs at time t ($CED(t)$), and the predictive value of an alarm at time t ($PV(t)$).

$$ARL^0: \quad E[t_A | D] \tag{9}$$

$$CED(t): \quad E[t_A - \tau | t_A \geqslant \tau, \tau = t] \tag{10}$$

$$PV(t): \quad P(C | t_A = t) \tag{11}$$

Note that the predictive value in (11) incorporates the time, since the predictive value is calculated for each time of alarm and is not necessarily the same for different alarm times. Like PPV in (8), the $PV(t)$ is a positive predicted value, but we used different notations to distinguish them.

In a hypothesis testing situation, the risk of Type I error is controlled by the size of the test, whereas in many on-line surveillance methods it is controlled by the ARL^0. A major drawback with a fixed size in on-line surveillance is that late changes are very difficult to detect, since the alarm limit must tend to infinity in order to keep the false alarm probability below for example 5 per cent, as demonstrated in [32]. In the next two sections, the OutbreakP method is evaluated for different situations. In Section 3, different outbreak detection methods are applied to Swedish tularaemia data, and in Section 4, the OutbreakP method is applied to Swedish influenza data.

3. DETECTION OF TULARAEMIA INFECTION OUTBREAKS

3.1. Tularaemia infection

The condition tularaemia is caused by an infection with the bacterium *Francisella tularensis*. According to the web site of the SMI, in Sweden the infection is mostly found among rodents, but it can also be transmitted to humans, for example by mosquitoes, water, or dust contaminated by urine or faeces from infected animals. In Sweden, tularaemia is a low-frequency disease. Large-scale outbreaks occurred in 1970, 1981, 2000, and 2003. According to the web site of the CDC, tularaemia is considered as a bioterrorism threat [33]. A weapon using airborne tularaemia would result in the outbreak of an acute, undifferentiated febrile illness with incipient pneumonia among other symptoms. Within an alert health system, an increased incidence should lead to the suspicion of intentionally caused tularaemia.

Tularaemia is an example of the twofold need for surveillance. Surveillance is not only needed for detecting natural outbreaks in order to allow health authorities to plan their actions, but also for detecting possible bioterrorist threats.

3.2. Outbreak detection

Tularaemia outbreaks in Sweden were monitored for the period 1998–2003 (see Figure 2) in [16]. We monitored the same data using the OutbreakP method and compared the results. Observe that

DOI: 10.1002/sim

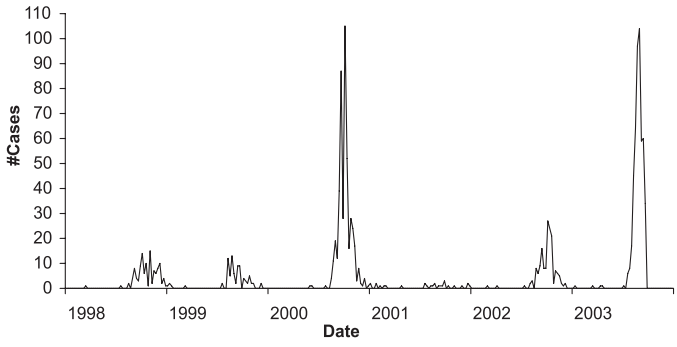

Figure 2. The number of tularaemia cases per week in Sweden for the period 1998–2003.

Table III. Evaluation of surveillance of tularaemia for the period 1998–2003.

Method	Sensitivity (per cent)	PPV (per cent)	Swiftness
OutbreakP	100	100	4
SPOTv2	100	100	3
EW	80	100	2
CuSum2	43	100	2
CuSum1	43	60	1

for the EW method data from earlier years were used according to automatic algorithm where a comparison with earlier years is made in order to detect unusual patterns.

In [16], several methods were evaluated. SPOTv2 is an earlier version of the method described in [17]. This method, which is used on a national level in Australia for monitoring salmonella infections, uses the Shewhart approach, see the alarm condition (3), with an advanced calculation of the baseline ($\mu^D(t)$). The EW method described in [19] is used routinely for many health-related events in England and Wales. The CuSum1 and CuSum2 methods described in the paper by Rolfhamre and Ekdahl are different variants of the CUSUM method by Page [22].

For the OutbreakP method, only data for the monitoring period 1998–2003 are used since that is enough to detect an outbreak according to (2)—that is increasing values. For the other methods, earlier data were also used according to the respective algorithms since the aim is to detect an unusual (e.g. seasonal) pattern.

In Table III, we compare the OutbreakP method with the results of other methods reported in [16]. The comparisons are made using the same measures as in [16]. These are retrospective measures for a fixed period, see the definitions in (6) and (8). Observe that the PPV used here is a measure for a fixed time (1998–2003) as opposed to PV defined in (11), which is used for on-line surveillance in the rest of the paper. The swiftness measure is the number of outbreaks (out of 5) for which each method was the first (by itself or simultaneously with another method) to give a warning. The conclusion from Table III is that the OutbreakP method compares well with the other methods. However, it must be kept in mind that the methods are designed to meet different aims and should rather be seen as complements than as competitors.

DOI: 10.1002/sim

Our method gives an alarm when the incidence increases—without regard of the cause. If the increase is due to the (almost) annual outbreak during summer, then the start of the seasonal outbreak (that varies much in time) is signalled. This might be useful for planning purposes of the medical authorities. The other methods are specifically constructed to detect an unusual seasonal pattern by signalling for incidences above the normal annual outbreaks. This will give a warning that something exceptional might have happened.

4. DETECTION OF INFLUENZA OUTBREAKS

The Swedish influenza incidence is measured by two weekly series, namely the number of laboratory diagnosed cases (LDI) and the percentage of patients showing influenza-like symptoms (%ILI) among all patients (denoted by #PAT) seen by sentinel physicians. For influenza, as opposed to tularaemia, we have no other study that can serve as a natural comparison with our method. A difficulty in evaluating methods for outbreak detection is that no official date exists as to the onset of the outbreak. Frisén and Andersson [15] evaluated the OutbreakP method in a simulation study, generating observations from a model that mimicked LDI data. In the same paper, the OutbreakP method was also applied to Swedish LDI data and was shown to have good properties.

In this paper we evaluate and apply the OutbreakP method to ILI data. In the simulation study presented in Section 4.1, the alarm limit is set to give an acceptable predictive value curve for the application to Swedish data. In Section 4.2, the OutbreakP method is applied to Swedish ILI data for six seasons. In order to illustrate the difficulties with prospective judgments as compared with retrospective ones, we present the results from an experiment with judgments by 26 medically trained individuals. These results are presented in Section 4.3.

4.1. Model for Swedish data on ILI and Monte Carlo evaluation

The ILI data are reported by sentinel physicians. Andersson *et al.* [3] discussed these data and their quality, and one conclusion was that ILI data should be interpreted with care, since they seem to be influenced by time-dependent effects, for example, physicians' inclination to send reports.

For the simulation study, we needed a model of how ILI data develop over time during the outbreak of the yearly epidemic. Data on ILI are reported both as the number of cases (#ILI) and as %ILI (the number of patients showing ILI in relation to the total number of patients). Swedish ILI data were analysed in [3] where it was concluded that a very simple model with a constant non-epidemic incidence and an exponential increase after the outbreak could be satisfactory as a first approximation. Another conclusion was that the distribution of the observations could be reasonably well described by a Poisson distribution. A Gaussian distribution is not appropriate during the onset of an outbreak with low incidence (see [14, 34]). The question of independence between observations close in time was also addressed in [3]. When modelling the autocorrelation of a process it is important to be aware that the estimated autocorrelation parameters do depend heavily on the assumptions made about the expected value of the process. In [3] it was concluded that the assumption of independent observations does work satisfactorily as a first approximation. The iid assumption is also used in [11, 12, 14, 34].

By using data for the first weeks of the latest 8 years, the incidence of the non-epidemic phase was roughly estimated to be 20 cases per week. For those outbreak detection methods where the observations are compared with a baseline (for example, Shewhart in (3)), the precision of this

DOI: 10.1002/sim

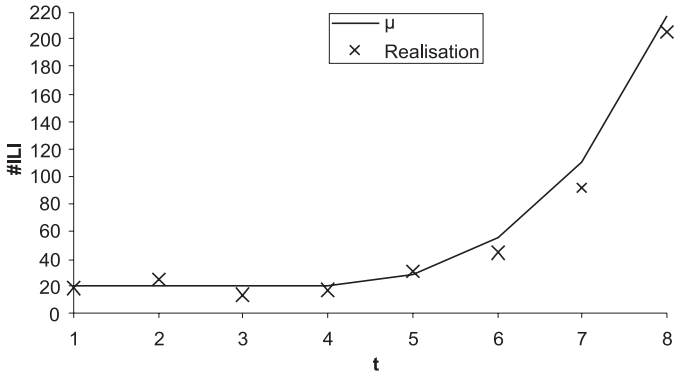

Figure 3. The $\mu(t)$ curve for $\tau=5$ representing the model for the expected numbers of patients showing influenza-like symptoms. Also an example of a simulated realization by the model is seen.

estimate is very important. However, here the estimate is used only to produce simulation results for situations that are close enough to the real situation to be of interest.

The exponential slope was estimated by the least-squares estimate of the logarithms of %ILI data for the season 2003/2004, which was a 'typical' season. Since we are using a surveillance system based on the Poisson likelihood, we needed the value of the 'number of ILI cases'. The actual #ILI values are not as representative for the progression of influenza as the %ILI values are, due to varying #PAT, as explained in the next section. In the model below we have transformed the %ILI numbers to #ILI by multiplying them with a constant.

Observations are generated according to the following model, for different values of τ (the time of the onset of the outbreak):

$$X(t)\sim\begin{cases}\text{Poi}(\mu_0), & t<\tau\\ \text{Poi}(\mu(t)), & t\geqslant\tau\end{cases}$$

where $\mu(t)=\exp(\beta_0+\beta_1(t-\tau+1))$ and $\mu_0=20$, $\beta_0=2.67$, and $\beta_1=0.68$. The Poisson distribution is denoted by Poi. One example of the $\mu(t)$ curve is given in Figure 3.

For each time τ of the onset of the outbreak, 1 000 000 replicates were used. For each replicate new observations were generated until the OutbreakP signalled an alarm. Two different limits were used; one warning limit and one alarm limit. The limits were chosen to give the alarms a high predicted value at different time points, as illustrated in Figure 4. The predictive value depends on the distribution of the outbreak time. Here it was assumed as a geometric distribution with an intensity $\nu=0.1$. This value was roughly estimated from Swedish data for 8 years by using the average time of the onset of the influenza epidemic.

4.2. Application of the OutbreakP method to Swedish ILI data

The results of the OutbreakP method are given below for six influenza seasons in Sweden. The ILI data of the present Swedish sentinel system are based on a varying total number of patients, as mentioned in Section 4.1. The Swedish system (where sentinel physicians report the percentage of patients with influenza-like illness during the past week) has the disadvantage of a low reporting

DOI: 10.1002/sim

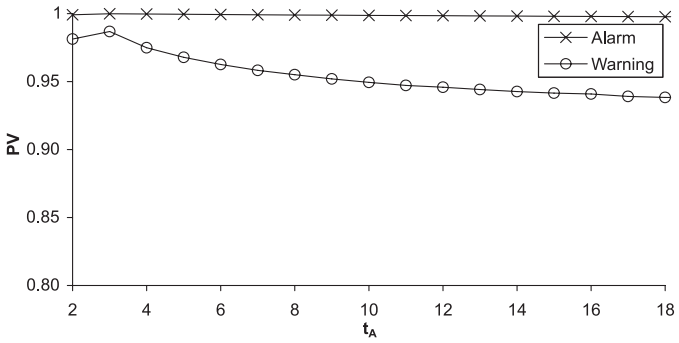

Figure 4. Predictive value for the OutbreakP method for $v=0.1$ and for the limits corresponding to the alarm limit and the warning limit, respectively. Results from a simulation study with a model resembling Swedish ILI data.

tendency in the beginning and at the end of the influenza season as well as during holidays, see [3]. Thus, #ILI is not representative for the true incidence and %ILI better represents the slope of the incidence. Great efforts are made to improve the reporting.

It is hard to base a stringent surveillance system on data with the selection bias described above. Because of the varying number of patients included in the sentinel system, it is not reasonable to use these incidences in surveillance. We demonstrate how the ILI data could be used if they were all based on a constant number of patients each week. In the surveillance situation described below, the OutbreakP method is applied to transformed values of %ILI, which represent values of #ILI for the ideal situation where the number of patients is constant at 15 000 each week. Hopefully, the current efforts will make this ideal situation a reality.

Figure 5 shows the OutbreakP method applied to the transformed %ILI values of six seasons. The same limits as in Figure 4 were used; the lower limit for early warnings and the higher one for more trustworthy alarms. Thus, the alarms had a predicted value of nearly one, unless the alarm had appeared at alarm time 2 (week 41), which none did. Owing to the assumptions and the quality of the data used these results must be interpreted with care, but they indicate a way to work with surveillance systems in outbreak detection.

4.3. Outbreak detection by subjective judgment

The outbreak is often easy to identify when studying the data retrospectively. In practise, however, the data become available sequentially, and the decisions have to be made in the same way (for example, a new decision on 'outbreak or not' each week). In order to compare the OutbreakP method with subjective judgments, we performed an experiment. We wanted to concentrate on the issue of subjective judgment without the complications of %ILI data based on a varying number of total patients. Therefore, this experiment was conducted using LDI data.

Twenty-six medically trained individuals participated in the experiment. The individuals consisted of all students in two lecture groups attending a course in medical statistics. We thus do not have a random sample from a well-defined population. However, the results will illustrate how subjective judgments can be compared with an objective method.

Each individual was planned to be given four sets of outbreak graphs (each set consisting of 10 graphs). Two sets displayed the situation of $\tau=5$ and the other two displayed the situation $\tau=9$.

DOI: 10.1002/sim

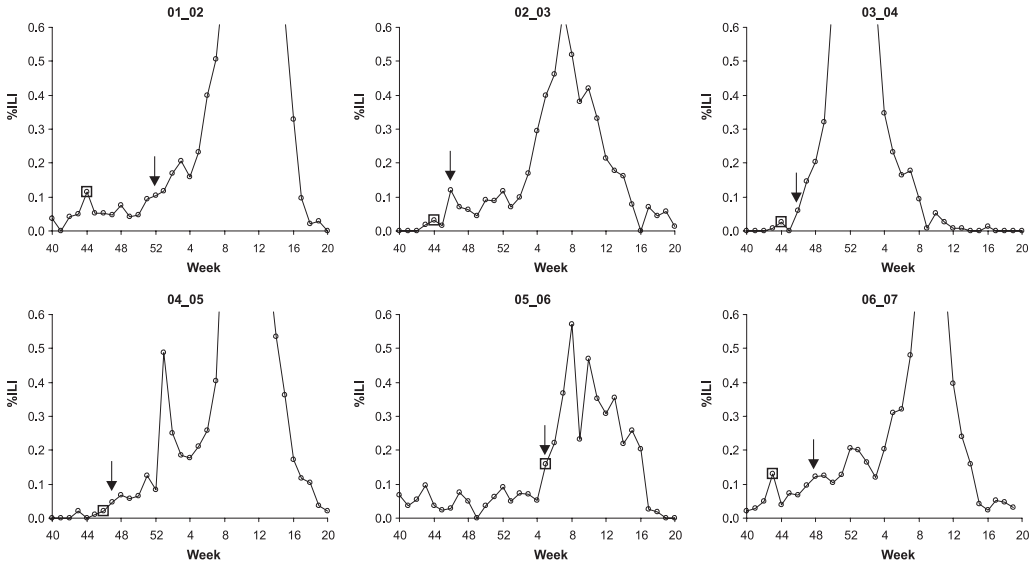

Figure 5. The %ILI data from the influenza periods 01/02 to 06/07 are displayed. The arrow marks the time of the alarm while the square marks the time of a warning when the OutbreakP method is used on the transformed data. The scale is chosen in order to set focus on the onset period. Thus, for some of the seasons, the peak is outside the graph.

Thus, two realizations of each value of τ were used. The order of the four sets was randomized. Owing to an administrative mistake, one individual only received three sets of graphs.

In order to be able to compare the results of the subjective judgments with those of the suggested method, we needed to know the true outbreak time (τ). Therefore, we produced the four outbreak sets by simulations. In [15] a simulation study for LDI data was performed, and here we used the same simulation model, namely

$$X(t) \sim \text{Poi}(\mu(t))$$

where

$$\mu(t) = \begin{cases} \mu_0, & t < \tau \\ \exp(\beta_0 + \beta_1 \cdot (t - \tau + 1)), & t \geqslant \tau \end{cases}$$

where $\mu_0 = 1$, $\beta_0 = 0.26$, $\beta_1 = 0.826$, and $\tau = \{5, 9\}$. The parameters were chosen to resemble the Swedish LDI data from the period 03 to 04.

The first six graphs of one situation are shown in Figure 6. Each subject looked at the graphs sequentially, i.e. one graph at a time, and as soon as he/she judged (from one of the graphs) that there had been an outbreak, this graph was marked.

The results of the experiment are presented in Figures 7 and 8. As can be seen in Figure 7, the distance between the distribution curves for $\tau = 5$ and $\tau = 9$ is a little smaller than what can be explained by the difference in the value of τ. This indicates that the subjective judgments tend to give an alarm too early for late outbreaks.

DOI: 10.1002/sim

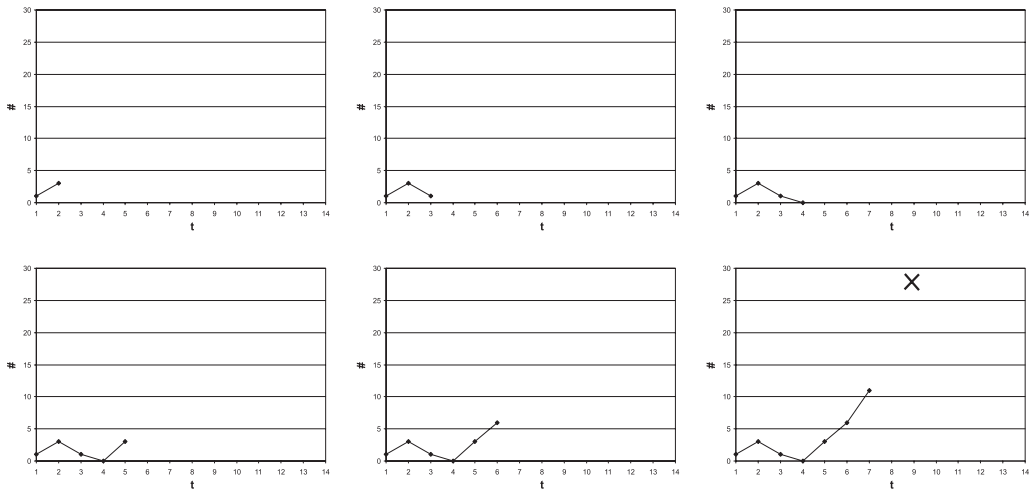

Figure 6. A set of graphs displaying the situation where $\tau=5$ for a simulated realization of LDI. The graph marked with X is the one where one of the subjects considered the outbreak to be located, i.e. $t_A=7$.

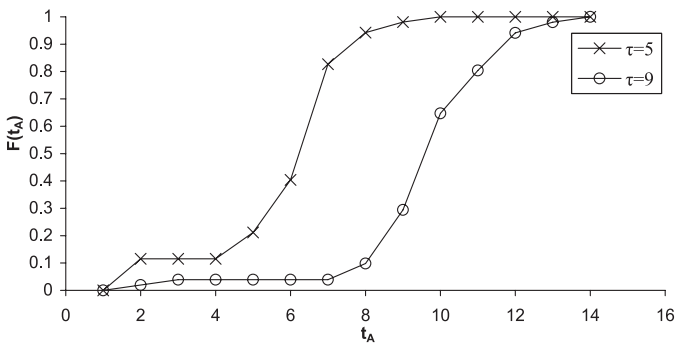

Figure 7. The distribution function for the alarm time by subjective judgment.

The subjects' judgments varied considerably also when they were given the same information. One alarm was given at time 2 and another at time 10 for $\tau=5$, for the same replicate. One advantage of using an algorithm is that the same information would always result in the same judgment. If we rely on subjective judgment, the result will depend heavily on which individual is making the judgment. One subject of the 26 was among the first to give an alarm in three of the cases and the second in the fourth.

The average delay times were 1.85 and 1.48 for $\tau=5$ and $\tau=9$, respectively. In the simulation study, the corresponding average delay times for the OutbreakP method were 1.58 and 1.48. The lower CED of the subjective judgment for $\tau=9$ can be due to random fluctuation or an expectation that a change should not occur too late.

DOI: 10.1002/sim

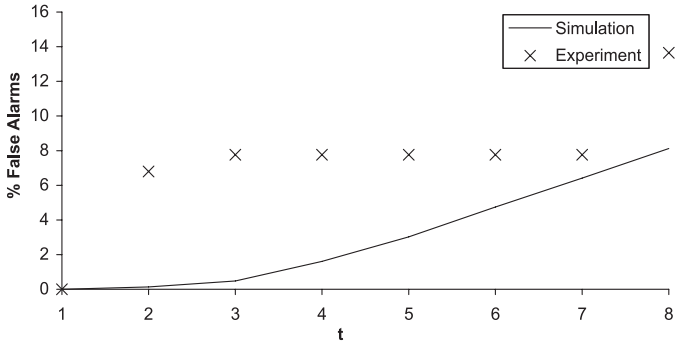

Figure 8. The false alarm probabilities $P(t_A \leqslant t|D)$ for the subjective judgment (experiment) and for the OutbreakP method (simulation).

In Figure 8, the false alarm distribution from the experiment is compared with that of the OutbreakP method. We can see that there is a higher cumulative probability of a false alarm in the experiment, compared with the algorithm. Since the delay is almost the same (or worse) for the subjective judgment as for the algorithm, we can conclude that the predicted values of the subjective judgments are lower.

5. DISCUSSION

An important question is what kind of outbreak we want to detect. We have found several explicit or implicit definitions of an outbreak. These different aims will correspond to different evaluations and different optimal methods. The methods should be seen as complements to each others. For each application, it is important to explicitly state the aim.

An important aim is to detect when the pattern (including seasonal variation) differs from that of earlier years. The method by Farrington *et al.* [19] has successfully been used in England Wales and other areas for this purpose.

Sometimes an identified set of related cases, such as a few cases arising from the same identified cause, is called an outbreak. An outbreak of this kind cannot be efficiently detected by a method for detecting a high incidence in the whole population but possibly by the detection of clustering of cases.

Some methods (see, for example, [35]) use an advanced model for the baseline. If the model is estimated from previous years' data, the outbreak is thus defined as a week of an unusually high incidence, compared with the incidence in the same week during earlier years' epidemics. In such a system we will get an early signal if the current year's influenza epidemic appears unusually early, but if it appears late, the signal will not be given until after the peak.

A general difficulty when estimating a baseline model that includes seasonality is that the seasonal effect may be hard to separate from the epidemic itself, since both often have a cycle of one year.

Some methods signal an outbreak when the incidence is high compared with a baseline estimated for non-epidemic periods. Errors in the estimated value can have a great impact, as demonstrated in [15].

DOI: 10.1002/sim

The semiparametric method used here detects outbreaks, which are defined by a monotonic increase after the constant level before the onset of the outbreak. Such outbreaks are of interest in connection with several diseases and syndromes. The method is based on the theory for optimal surveillance. A user-friendly computer program is available. When seasonal effects or other explaining variables are present, residuals from a model that incorporates seasonality can be monitored. However, estimation of the seasonal component from the same data that are being monitored does affect all surveillance methods and the effect should be studied.

When evaluating methods for on-line monitoring it is important to use measures that incorporate the time issue, i.e. the fact that there are repeated decisions, not just one decision as in hypothesis testing.

The semiparametric method was applied to Swedish data on tularaemia, which had previously been analysed in [16]. We used the same data and the same (conventional) evaluation measures as in this study. The OutbreakP method came out favourably.

The OutbreakP method was further evaluated in a simulation study, where data were generated from a model mimicking the behaviour of Swedish data on influenza-like illness. In the construction of the simulation model, we also discussed how to handle data quality problems. Here, we used evaluation measures, which are better suited for on-line surveillance than the conventional ones.

We applied our method to Swedish ILI data from six seasons with good results. On-line surveillance, where only a limited amount of information is at hand at each decision time, is much more demanding the retrospective identification of the outbreaks. Thus, we made an experiment with subjective judgments by medically trained individuals. The subjective method was less efficient than the OutbreakP method. However, the main disadvantage turned out to be the large variation between judges.

When data from different sources are available, multivariate surveillance should be applied. This can be the case at detection of outbreaks of influenza based on data from different regions or at detection of bioterrorism based on different kinds of data. The two simplest approaches of multivariate surveillance are a reduction to a suitable univariate statistic or a parallel surveillance with due concern of the multiplicity. However, there are also other approaches [36]. The dependencies play an important role for the choice of the optimal approach [37].

Hopefully, the suggested method will add to the toolbox for outbreak detection.

ACKNOWLEDGEMENTS

We thank the referee and the Associate Editor for valuable suggestions. Kjell Pettersson gave constructive comments. The data were made available to us by the Swedish Institute for Infectious Disease Control, and we are grateful for discussions about the aims and the data quality. The research was supported by the Swedish Emergency Management Agency (grant 0314/206).

REFERENCES

1. Doyle A, Bonmarin I, Lévy-Bruhl D, Le Strat Y, Desenclos J-C. Influenza pandemic preparedness in France: modelling the impact of interventions. *Journal of Epidemiology and Community Health* 2006; **60**:399–404. DOI: 10.1136/jech.2005.034082.
2. Hall IM, Gani R, Hughes HE, Leach S. Real-time epidemic forecasting for pandemic influenza. *Epidemiology and Infection* 2006; **135**:372–385. DOI: 10.1017/S0950268806007084.
3. Andersson E, Bock D, Frisén M. Modeling influenza incidence for the purpose of on-line monitoring. *Statistical Methods in Medical Research* 2008; **17**:421–438. DOI: 10.1177/0962280206078986.

4. Andersson E, Kuhlmann-Berenzon S, Linde A, Schiöler L, Rubinova S, Frisén M. Predictions by early indicators of the time and height of yearly influenza outbreaks in Sweden. *Scandinavian Journal of Public Health* 2008; **36**:475–482. DOI: 10.1177/1403494808089566.
5. Linde A, Brytting M, Johansson M, Wiman Å, Högberg L, Ekdahl K. *Annual Report September 2003–August 2004*. The National Influenza Reference Center, Swedish Institute for Infectious Disease Control, Stockholm, 2004.
6. Kulldorff M. Prospective time periodic geographical disease surveillance using a scan statistic. *Journal of the Royal Statistical Society*, *Series A* 2001; **164**:61–72. DOI: 10.1111/1467-985X.00186.
7. Sonesson C. A CUSUM framework for detection of space–time disease clusters using scan statistics. *Statistics in Medicine* 2007; **26**:4770–4789. DOI: 10.1002/sim.2898.
8. Rosychuk RJ, Huston C, Prasad NGN. Spatial event cluster detection using a compound Poisson distribution. *Biometrics* 2006; **62**:465–470. DOI: 10.1111/j.1541-0420.2005.00503.x.
9. Bock D, Pettersson K. Exploratory analysis of spatial aspects on the Swedish influenza data. *Smittskyddsinstitutets Rapportserie 3:2006*, Swedish Institute for Infectious Disease Control, Stockholm, 2006.
10. Lawson AB, Kleinman K (eds). *Spatial and Syndromic Surveillance for Public Health*. Wiley: New York, 2005.
11. Stroup DF, Thacker SB, Herndon JL. Application of multiple time-series analysis to the estimation of pneumonia and influenza mortality by age 1962–1983. *Statistics in Medicine* 1988; **7**:1045–1059. DOI: 10.1002/sim.4780071006.
12. Serfling R. Methods for current statistical analysis of excess pneumonia–influenza deaths. *Public Health Reports* 1963; **78**:494–506.
13. Quenel P, Dab W, Hannoun C, Cohen JM. Sensitivity specificity and predictive values of health service based indicators for the surveillance of influenza A epidemics. *International Journal of Epidemiology* 1994; **23**:849–855. DOI: 10.1093/ije/23.4.849.
14. Le Strat Y, Carrat F. Monitoring epidemiologic surveillance data using hidden Markov models. *Statistics in Medicine* 1999; **18**:3463–3478. DOI: 10.1002/(SICI)1097-0258(19991230)18:24<3463::AID-SIM409>3.0.CO;2-I.
15. Frisén M, Andersson E. Semiparametric surveillance of outbreaks. *Research Report 2007:11*, Statistical Research Unit, Department of Economics, Göteborg University, Sweden, 2007. Available from: http://hdl.handle.net/2077/10527.
16. Rolfhamre P, Ekdahl K. An evaluation and comparison of three commonly used statistical models for automatic detection of outbreaks in epidemiological data of communicable diseases. *Epidemiology and Infection* 2006; **134**:863–871. DOI: 10.1017/S095026880500573X.
17. Stern L, Lightfoot D. Automated outbreak detection: a quantitative retrospective analysis. *Epidemiology and Infection* 1999; **122**:103–110. DOI: 10.1017/S0950268898001939.
18. Farrington CP, Beale AD. Computer-aided detection of temporal clusters of organisms reported to the Communicable Disease Surveillance Centre. *Communicable Disease Report* 1993; **3**:R78–R82.
19. Farrington CP, Andrews NJ, Beal AD, Catchpole MA. A statistical algorithm for the early detection of outbreaks of infectious disease. *Journal of the Royal Statistical Society*, *Series A* 1996; **159**:547–563.
20. Farrington CP, Beale AD. The detection of outbreaks of infectious disease. In *Geomed 97*, Gierl L, Cliff A, Valleron AJ, Farrington CP, Bull M (eds). Teubner Verlag: Leipzig, 1998.
21. Farrington CP, Andrews NJ. Outbreak detection: application to infectious disease surveillance. In *Monitoring the Health of Populations*, Brookmeyer R, Stroup DF (eds). Oxford University Press: Oxford, 2004.
22. Page ES. Continuous inspection schemes. *Biometrika* 1954; **41**:100–114. DOI: 10.1093/biomet/41.1-2.100.
23. Shewhart WA. *Economic Control of Quality of Manufactured Product*. MacMillan and Co.: London, 1931.
24. Shiryaev AN. On optimum methods in quickest detection problems. *Theory of Probability and its Applications* 1963; **8**:22–46. DOI: 10.1137/1108002.
25. Roberts SW. A comparison of some control chart procedures. *Technometrics* 1966; **8**:411–430.
26. Andersson E, Bock D, Frisén M. Statistical surveillance of cyclical processes. Detection of turning points in business cycles. *Journal of Forecasting* 2005; **24**:465–490. DOI: 10.1002/for.966.
27. Frisén M. Statistical surveillance of business cycles. *Research Report 1994:1*, Department of Statistics, Göteborg University, Sweden, 1994.
28. Frisén M. Unimodal regression. *The Statistician* 1986; **35**:479–485.
29. Bock D, Andersson E, Frisén M. Statistical surveillance of epidemics: peak detection of influenza in Sweden. *Biometrical Journal* 2008; **50**:71–85. DOI: 10.1002/bimj.200610362.
30. Frisén M, Andersson E, Pettersson K. Semiparametric estimation of outbreak regression. *Research Report 2007:13*, Statistical Research Unit, Department of Economics, Göteborg University, Sweden, 2007. Available from: http://hdl.handle.net/2077/10526.

31. Robertson T, Wright FT, Dykstra RL. *Order Restricted Statistical Inference*. Wiley: Chichester, 1988.
32. Bock D. Aspects on the control of false alarms in statistical surveillance and the impact on the return of financial decision systems. *Journal of Applied Statistics* 2008; **35**:213–227. DOI: 10.1080/02664760701775431.
33. Chang M, Glynn MK, Groseclose SL. Endemic, notifiable bioterrorism-related diseases, United States, 1992–1999. *Emerging Infectious Diseases* 2003; **9**:556–564.
34. Rath TM, Carreras M, Sebastiani P. Automated detection of influenza epidemics with hidden Markov models. In *Advances in Intelligent Data Analysis V*, Berthold MR, Lenz H-J, Bradley E, Kruse R, Borgelt C (eds). Springer: Berlin, Germany, 2003.
35. Sebastiani P, Mandl KD, Szolovits P, Kohane IS, Ramoni MF. A Bayesian dynamic model for influenza surveillance. *Statistics in Medicine* 2006; **25**:1803–1816. DOI: 10.1002/sim.2566.
36. Frisén M. Principles for multivariate surveillance. In *Frontiers in Statistical Quality Control*, Lenz H-J, Wilrich P-T (eds). 2008; in press.
37. Andersson E. Effect of dependency in systems for multivariate surveillance. *Research Report 2007:1*, Statistical Research Unit, University of Gothenburg, 2007.

Paper II

Journal of Applied Statistics

Taylor & Francis
Taylor & Francis Group

Evaluation of multivariate surveillance

Marianne Frisén[a]*, Eva Andersson[a,b] and Linus Schiöler[a]

[a]*Statistical Research Unit, Department of Economics, University of Gothenburg, Gothenburg, SE 40530, Sweden;* [b]*Department of Occupational and Environmental Medicine, Sahlgrenska University Hospital and Sahlgrenska Academy, University of Gothenburg, Gothenburg, SE 40530, Sweden*

(*Received 18 April 2009; final version received 28 July 2009*)

Multivariate surveillance is of interest in many areas such as industrial production, bioterrorism detection, spatial surveillance, and financial transaction strategies. Some of the suggested approaches to multivariate surveillance have been multivariate counterparts to the univariate Shewhart, EWMA, and CUSUM methods. Our emphasis is on the special challenges of evaluating multivariate surveillance methods. Some new measures are suggested and the properties of several measures are demonstrated by applications to various situations. It is demonstrated that zero-state and steady-state ARL, which are widely used in univariate surveillance, should be used with care in multivariate surveillance.

Keywords: average run length; EWMA; false alarms; FDR; performance metrics; predictive value; steady state; zero state

1. Introduction

In many situations, there are reasons to continuously observe a process in order to detect an important change in the process as soon as possible after the change has occurred. Multivariate surveillance typically concerns several variables. However, it is also of interest when there is only one process, but several characteristics of that process may change. Examples are processes where both the mean and the variance may change (treated in [19]) or changes in several aspects of one autoregressive time series take place, as treated in [3].

The first suggestion of modern control charts [33] was widely utilized by industry. The monitoring of several processes is often of interest. Multivariate problems for the assembly process of the Saab automobile were described in [37]. In the food industry, different raw materials and several process steps are used, and in [32] it is suggested that these be analyzed in order to assure the quality of the final product. During the last years there have been an increased need for, and interest in, continuous monitoring in many areas apart from industrial production. After the 9/11 attack the interest in surveillance methodology increased notably in the US, and new types of data are now being collected to get early signals of bioterrorism. By monitoring several data series different aspects can be covered, and thus multivariate surveillance techniques are

*Corresponding author. Email: marianne.frisen@statistics.gu.se

ISSN 0266-4763 print/ISSN 1360-0532 online

needed. Rolka et al. [28] contains an overview of the research needs for bioterrorism surveillance using multiple data streams. Spatial surveillance is another example of multivariate surveillance, since several locations are involved. A relatively new area for multivariate surveillance is financial decision strategies in situations where a portfolio contains several assets [13,25].

General reviews on multivariate surveillance methods are made for example in [4,6,9,21,31,36,37].

Multivariate surveillance can have different aims. Sometimes, the aim is to identify the parameters that have changed. However, this is naturally preceded by the detection of a change in any of the parameters. Here, we concentrate on the detection of the first change.

2. Notations and specifications

We will denote the multivariate process under surveillance by a p-variate vector, $\mathbf{Y}(\mathrm{t}) = \{Y_1(t), Y_2(t), \ldots, Y_p(t)\}$. The components of the vector may be, for example, a measurement on p different processes. The distribution of the p-variate variable $\mathbf{Y}(t)$ might be characterized by the mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}_{\mathbf{Y}}$. The aim is to detect the change from one state – for example that the assembled product works well – to another – that some component is defective so that the product does not work. We aim to detect the change as soon as possible after it has occurred in order to give warning and take corrective action. At decision time s, we base the decision on the available information $\mathbf{Y}^{\mathrm{s}} = \{\mathbf{Y}(1), \mathbf{Y}(2), \ldots, \mathbf{Y}(\mathrm{s})\}$ to form an alarm statistic. This is used to define the alarm rule.

In the multivariate situation, we observe p processes that can change at different times $\tau_1, \ldots, \tau_p$. Here the aim is to detect frequently the first time that the process is no longer in control – that is, we want to make inference about $\tau_{\min} = \min\{\tau_1, \ldots, \tau_p\}$. If there is no change at all, we denote this by "$\tau_{\min} = \infty$".

3. Surveillance methods

We will discuss different evaluation metrics for multivariate surveillance. The discussion is supported by results where commonly used methods are evaluated by the metrics. The evaluation measures will reveal the principal differences between the approaches for multivariate surveillance.

3.1 *General approaches*

3.1.1 *Dimension reduction*

One approach for handling multivariate surveillance problems is to reduce the p-variate vector at each time point into a single statistic and then use a system for univariate surveillance based on this statistic. One may simply use the sum or another linear combination of the variables. When we want to derive an optimal method, we must specify the type of change that we want the method to detect. One way to focus attention is to consider some type of dimension reduction transformation as in [14,15,29]. In [30], this is done with specific respect to the special and common causes of variation. Sometimes a sufficient reduction can be found as in [38], where it is proved that when the changes occur simultaneously, it is possible to find a sufficient reduction to a univariate surveillance problem. For the exponential family with the same shift and dispersion parameter and independence between the processes, conditional on change times, the sufficient statistic for each time t is the sum of the observations $Y_1(t), \ldots, Y_p(t)$. For some situations where the changes occur with known lags, it is also possible to find a sufficient reduction, see [17]. When

a sufficient reduction is found, optimal methods can be derived. In many situations, however, it is not possible to find a sufficient reduction.

3.1.2 *Parallel surveillance*

The approach with parallel systems means that one starts with a univariate surveillance method for each variable. The most common way to combine the information from the p univariate methods is to signal an alarm if any of the univariate methods gives an alarm. This approach is called combined univariate methods or parallel methods.

3.2 *Specific methods and situations*

3.2.1 *Example*

We will illustrate the suggested measures and their properties by applying them in a number of different situations and for different methods. We will concentrate on the way in which the time of the changes influences the properties, and therefore a very simple example with two processes will be used. Our model contains two normally distributed variables, Y_1 and Y_2, which possibly have shifts in the expected value at possibly different time points. In order to focus on the effect of different change times we use equal shift sizes. The two processes, Y_1 and Y_2, are assumed to be independent (conditional on the change times).

$$Y_1(t) \sim \begin{cases} N(0,\,1) & t < \tau_1 \\ N(1,\,1) & t \geq \tau_1 \end{cases}$$

$$Y_2(t) \sim \begin{cases} N(0,\,1) & t < \tau_2 \\ N(1,\,1) & t \geq \tau_2. \end{cases}$$

The alarm times for different methods were determined by Monte Carlo simulations with at least 10,000,000 replicates for each situation.

3.2.2 *Specific methods*

Multivariate methods are usually extensions of common univariate methods. The univariate technique used here is the EWMA method, since it is commonly used in multivariate situations also. By this method we use exponential weights for the observations, giving more weight to recent observations and less to old ones. The statistic at decision time s, of the EWMA method for univariate surveillance of a variable Y with target value zero is

$$\mathrm{Z}_s = \lambda(1-\lambda)^s \sum_{t=1}^{s} (1-\lambda)^{-t} Y(t),$$

where $0 < \lambda \leq 1$. Regarding the variance of the EWMA statistic there are two versions: the exact and the asymptotic variance, and we will use the asymptotic version as recommended in [35]. This means that the statistic is compared with a constant alarm limit.

The optimal value of the parameter λ has drawn much attention. A formula for the optimal value was derived in [9] and explicitly given in [11] as $\lambda^* = 1 - \exp(-\mu^2/2)/(1-\nu)$, where μ is the shift size (here $\mu = 1$) and $\nu = P(\tau = t | \tau \geq t)$ denote how often changes are prone to occur. Here we choose the value $\lambda = 0.35$ that will give an approximately optimal method for a wide range of ν.

We will compare results from (i) the EWMA method applied to a reduction of data to a univariate statistic at each time, (ii) a system based on two parallel EMWA methods, and (iii) the EWMA method applied to the univariate process that changes first.

As an example of the reduction approach, we reduce the bivariate variable (Y_1, Y_2) to a univariate statistic, here chosen to

$$R(t) = \frac{Y_1(t) + Y_2(t)}{2}.$$

Then the EWMA method is applied to the variable $R(t)$ (with the variance $\sigma_R^2 = 0.5$). The time of alarm for the reduction approach, t_{AR}, is the first time when the EWMA statistic exceeds a constant alarm limit.

The parallel approach means that the EWMA method is applied to $Y_1(t)$ and $Y_2(t)$ separately. The time of alarm for Y_1, t_{A1}, is the first time when EWMA_{Y1} exceeds a constant alarm limit (correspondingly for Y_2). The time of alarm for the parallel approach is the first of either of the alarm times ($t_{AP} = \min[t_{A1}, t_{A2}]$.

For comparison, we also have the results from the EWMA method applied to only one process. This corresponds to the situation when there is prior knowledge about which process will change first, and therefore it is efficient to monitor only this one.

The alarm limits are set in order to give each of the systems the same false alarm property.

4. Evaluation metrics

Timeliness in detection is of extreme interest in surveillance, and hence there is a need for other evaluation measures than those traditionally used in hypothesis testing.

4.1 *False alarms*

In a univariate setting, the most commonly used measure is $\text{ARL}^0 = E[t_A|\tau = \infty]$ where t_A is the time of alarm. This is naturally generalized for multivariate surveillance as $E[t_A|\tau_{\min} = \infty] = E[t_A|\tau_1 = \infty, \ldots, \tau_p = \infty]$ where t_A is the time of the general alarm for the multivariate situation. The median run length, MRL^0, can be used instead of the expected value with the same generalization as for average run length, ARL^0. In the simulations below, the alarm limits are set so that each of the systems has an MRL^0 equal to 100.

In univariate theoretical work the false alarm probability, $\text{PFA} = P(t_A < \tau)$, is commonly used. This is naturally generalized for multivariate surveillance as

$\text{PFA} = P(t_A < \tau_{\min}) = \sum_{i=1}^{\infty}(P(t_A < \tau_{\min}\,|\tau_{\min} = i\,)P(\tau_{\min} = i))$. It can also be expressed as $\text{PFA} = P(t_A < \tau_j)P(\tau_{\min} = \tau_j)$.

Note that the distribution of $\tau_{\min}$ (through the distribution of the change point distributions of all variables) is included in the suggested multivariate PFA expression.

In hypothesis testing with multiple comparisons, it is important to control the probability of false rejection (an overview of important methods is given in [16]). For the situation when several drugs are tested against one standard, the family-wise error rate is relevant. This is the probability of any false rejection for the family of sub-hypotheses. For another situation, for example when several aspects of a single drug are tested, the false discovery rate (FDR), suggested in [5] may be more relevant. This is the proportion of rejections that are false. Recently, FDR has been suggested for surveillance problems for example in [28]. Surveillance where we make more than one decision, differs from hypothesis testing in that methods with high detection ability have a false alarm rate that tends to one (as time tends to infinity), see for example [7]. If one tries to avoid this, by letting the alarm limit tend to infinity, it will harm the ability to detect late changes. Thus, false alarms are not regarded in the same way in surveillance as in hypothesis testing. In

hypotheses testing, it is important to control the probability of a false alarm. Since false alarms are unavoidable in surveillance, it is the frequency of false alarms in time which is important. Thus, in surveillance, the FDR is not needed. On the other hand, the ARL^0 of the multivariate procedure, as suggested above, might be easily interpreted as the expected time until a false alarm. In addition to this, the FDR measure is difficult to use in surveillance, since it is based on a probability that is not constant. There are different suggestions for solving this problem: In [23], a fairly short period of time is monitored and only the properties of the early part of the run length is used.

4.2 *Delay*

4.2.1 *Delay as a function of the time of the change*

We start by recapturing the univariate case where the expected delay for a specific value of τ is

$$\mathrm{ED}(\tau) = E\{\max(0, t_\mathrm{A} - \tau)\},$$

or, if τ is stochastic, the average delay over the distribution of τ

$$\mathrm{ED} = E\{\mathrm{ED}(\tau)\}.$$

This average is the base for the ED optimality, which is closely related to the utility functions suggested in [34], and sometimes called a Bayesian measure since it depends on the distribution of τ, which for some applications is naturally regarded as a parameter, and for others, as a stochastic variable.

Since $\mathrm{ED}(\tau)$ for most methods tends to zero (because of the false alarms when τ tends to infinity), it is useful to study the delay conditional on no alarms before τ. For a specific value of τ, the conditional expected delay, CED, is

$$\mathrm{CED}(\tau) = E[t_\mathrm{A} - \tau \,|t_\mathrm{A} \geq \tau\,].$$

The first use of the term CED and a calculation for a specific value of τ different from 1 and ∞ seems to be in [41]. In [1,12], the CED was used as a function of τ, and in [9,11] it was strongly advocated that the whole CED curve be studied. In [20], the dependency on τ is avoided by using the least favorable value of τ. The asymptotic measure is another example of how the value of τ can be avoided. The CED has been a component in many measures, but often in a way that avoids the dependency on τ.

In the multivariate case, the $\mathrm{ED}(\tau_1, \ldots, \tau_p)$ and $\mathrm{CED}(\tau_1, \ldots, \tau_p)$ depend on the vector $\{\tau_1, \ldots, \tau_p\}$, and ED depends on the multivariate distribution of $(\tau_1, \ldots, \tau_p)$. In [2], the following delay measure (for a situation where $p = 2$) was suggested and the dependency on $\tau_{\min}$ was demonstrated

$$\mathrm{CED}(\tau_1, \tau_2, \ldots, \tau_p) = E[t_\mathrm{A} - \tau_{\min} \,|t_\mathrm{A} \geq \tau_{\min}\,].$$

This delay measure depends on all the change points. However, there is often some relation between the change times which simplifies the picture. In Figures 1 and 2, we will use the multivariate CED to demonstrate principal differences between methods for some typical situations with special relations between the change times. The CED for different relations between the change points is presented for the reduction approach in Figure 1 and for the parallel approach in Figure 2.

By comparison between Figures 1 and 2, we can see that the general results in [38] mentioned in Section 3.1.1 hold here: the reduction approach is superior (gives shorter delay) when all processes change at the same time. The CED curves differ considerably for different relations between the values of the change times. This supports the need for the generalized CED measure.

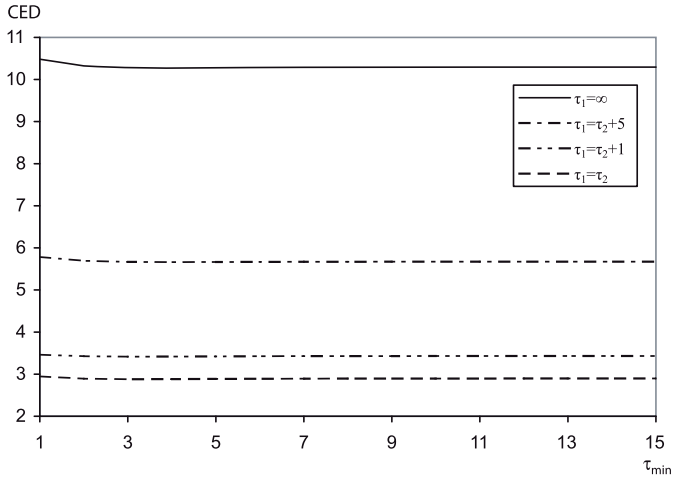

Figure 1. CED(τ_1, τ_2) *vs.* $\tau_{\min}$ for different relations between the τ values, presented for the reduction approach.

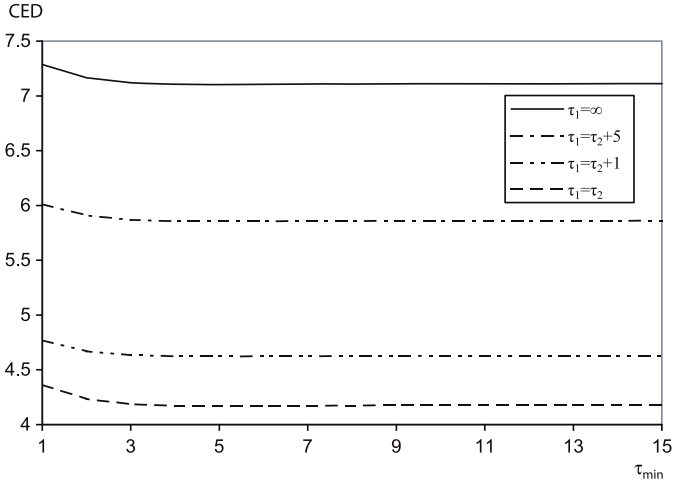

Figure 2. CED(τ_1, τ_2) *vs.* $\tau_{\min}$ for different relations between the τ values, presented for the parallel approach.

Sometimes the time available for action is limited. In such situations, it is important to use a surveillance system with high detection ability within the limited time available. This property can be measured by the probability of successful detection, which was suggested by [8]. It measures the probability that an alarm is called within d time points. In the multivariate case it can be defined as

$$\mathrm{PSD}(d, \tau_1 \ldots, \tau_p) = P(t_A - \tau_{\min} \leq d | t_A \geq \tau_{\min}),$$

as in [10].

The PSD measure is a function of both the change times $(\tau_1, \ldots, \tau_p)$ and the length of the interval in which the detection is defined as successful (d). In [39], it is suggested that the PSD be calculated as a function of only d and $\tau_{\min}$, by expressing PSD as an expected value for (stochastic) change points $(\tau_1, \ldots, \tau_p)$. The PSD can be used to describe the detection ability of a method

and compare it with that of other methods. PSD can also be calculated and compared for different values of d, as is done in [23] in connection with the use of the FDR. If we expect sudden, major changes, we may want a method with high detection ability (a high PSD for a small d). In a situation where we expect small changes, the long-term detection ability (a high PSD for a large d) may be more important. Thus it is essential to consider what kind of change one wants to detect at different time points. In Figure 3, we examine the PSD for the parallel approach, for two different cases of relations between the change points. With the parallel approach, it is easier to quickly detect simultaneous changes than changes quickly with a time lag. The PSD will tend to one for both cases when d increases.

4.2.2 *Zero-state ARL*

One measure of detection ability is the ARL, given that the change occurs immediately ($\tau = 1$). This is widely used in univariate surveillance and often named zero-state ARL or ARL^1. In univariate surveillance, the ARL^1 has a simple relation to the delay, namely $ARL^1 = CED(1) + 1$. This demonstrates that only $\tau = 1$ is considered. It may also be important to consider other change times, since the delay and detection ability of many methods depend on when the change occurs (i.e. depend on τ). To consider only $\tau = 1$ in the univariate case is a limitation, and the univariate ARL^1 is criticized as a formal optimality criterion, for example by [9].

Zero-state ARL is the most commonly used evaluation measure in the multivariate case also. However, it is seldom explicitly defined. One possibility is to define the multivariate zero-state ARL as $E[t_A|\tau_{min} = 1]$. However, as seen in Figure 2, the values of CED for $\tau_{min} = 1$ vary a lot for different relations between the values of τ_{min} and the change times of the other processes. Thus, there is no unique zero-state ARL with the definition $E[t_A|\tau_{min} = 1]$. Another possibility is to define the multivariate zero-state ARL as $E[t_A|\tau_1 = \tau_2 = \cdots = \tau_p = 1]$. This is probably the definition implicit in most publications. Here, it is assumed that all processes change at the same time. It was demonstrated by [38] that a sufficient reduction to a univariate problem exists when all processes change at the same time. Thus, when $\tau_1 = \cdots = \tau_p = 1$, a reduction to a univariate surveillance statistic is the proper procedure by the sufficiency principle, which means

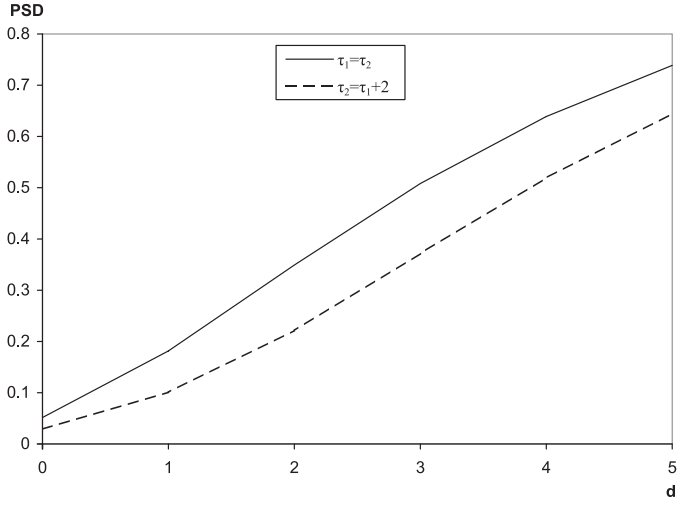

Figure 3. PSD(τ_1, τ_2) *vs.* d for different relations between the τ values ($\tau_1 = \tau_2$ and $\tau_1 + 2 = \tau_2$) presented for the parallel approach when $\tau_{min} = 3$.

that we have a univariate situation. Zero-state ARL is thus questionable as a formal measure for comparing methods for genuinely multivariate problems.

4.2.3 *Steady-state delay*

Already Roberts [27] suggested the use of the limit of CED as τ tends to infinity (even though he used τ=8 in the numerical comparisons). Here, this will be called the steady-state CED, CED_{SS}, and be defined as

$$\text{CED}_{\text{SS}} = \lim_{\tau\to\infty} CED(\tau).$$

This steady-state delay is closely related to steady-state ARL (often denoted by SS ARL or ARL_{SS}) which is defined in [22] as "the time from the change to the signal... using the steady state distribution" or more specifically in [18] as

$$\lim_{\tau\to\infty} E[t_{\text{A}} - \tau + 1 \,|t_{\text{A}} \geq \tau\,].$$

Here we see that $\text{ARL}_{\text{SS}} = \text{CED}_{\text{SS}} + 1$. This corresponds to the relation $\text{ARL}^1 = \text{CED}(1) + 1$.

Evaluations of multivariate methods by asymptotic measures are often made by the same measures as are used for univariate methods. For example, Lu and Reynolds [22] used "the steady state average delay time" and Reynolds and Kim [26] "the steady-state average time to signal." However, the correspondence to the univariate CED_{SS} is not without problems. The multivariate CED depends on several τ values and so does the multivariate steady-state CED, as seen in Figure 2. There is thus no unique steady-state CED (or steady-state ARL) that could characterize a method. The example supports this general result. Often only the situation $\tau_1 = \tau_2 = \cdots = \tau_p$ is considered. In that case we have

$$\text{CED}(\tau_1 = \tau_2 = \cdots = \tau_p = t) \text{ as } t \to \infty.$$

For equal change points we have a unique delay value for each method. However, this is another example of the situation where univariate surveillance can be used instead of multivariate surveillance since there is a sufficient reduction to univariate surveillance. This is confirmed by Figures 1 and 2, where we saw that the best method is based on the reduction to a univariate statistic. For other situations than simultaneous changes there is no simple asymptotic CED, as is seen in Figures 1 and 2. Even though all the τ values tend to infinity, it also matters how they do this. There is no simple asymptotic measure for the multivariate case. Instead, one has to specify how the times of the change points are related when they tend to infinity.

4.3 *Predictive value*

The predictive value suggested in [8] is defined as

$$\text{PV(t)} = P(C(t)|t_{\text{A}} = t) = \frac{\text{PMA}(t)}{\text{PMA}(t) + \text{PFA}(t)},$$

where C(t) is the change to be detected at decision time t, PMA the probability of a motivated alarm, and PFA the probability of a false alarm. Thus, $\text{PMA}(t) = P(C(t)|t_{\text{A}} = t)$ and $\text{PFA} = P(D(t)|t_{\text{A}} = t)$, where $D(t)$ is the in-control situation. The exact specification of $C(t)$ depends on the application. The most usual specification is $C(t) = \{\tau \leq t\}$ but others also can be relevant as seen below.

In a univariate setting with $C(t) = \{\tau \leq t\}$ and $D(t) = \{\tau > t\}$ this is

$$\mathrm{PV}(t) = P(\tau \leq t | t_{\mathrm{A}} = t) = \frac{\sum_{i=1}^{t} (P(t_{\mathrm{A}} = t | \tau = i) P(\tau = i))}{\sum_{i=1}^{t} (P(t_{\mathrm{A}} = t | \tau = i) P(\tau = i)) + P(t_{\mathrm{A}} = t | \tau > t) P(\tau > t)}.$$

In a multivariate setting, we generalize this with $C(t) = \{\tau_{\min} \leq t\}$ and $D(t) = \{\tau_{\min} > t\}$ to

$$\begin{aligned}\mathrm{PV}(t) &= P(\tau_{\min} \leq t | t_{\mathrm{A}} = t) \\ &= \frac{\sum_{i=1}^{t} (P(t_{\mathrm{A}} = t \,| \tau_{\min} = i\,) P(\tau_{\min} = i))}{\sum_{i=1}^{t} (P(t_{\mathrm{A}} = t \,| \tau_{\min} = i\,) P(\tau_{\min} = i)) + P(t_{\mathrm{A}} = t \,| \tau_{\min} > t\,) P(\tau_{\min} > t)}.\end{aligned}$$

For the case of two variables, Y_1 and Y_2, we have that the probabilities of a motivated and a false alarm, respectively, are

$$\begin{aligned}\mathrm{PMA}(t) = &\sum_{i=1}^{t} \sum_{j=1}^{t} (P(t_{\mathrm{A}} = t | \tau_1 = i, \tau_2 = j) P(\tau_1 = i, \tau_2 = j)) \\ &+ \sum_{i=1}^{t} (P(t_{\mathrm{A}} = t | \tau_1 = i, \tau_2 > t) P(\tau_1 = i, \tau_2 > t)) \\ &+ \sum_{j=1}^{t} (P(t_{\mathrm{A}} = t | \tau_1 > t, \tau_2 = j) P(\tau_1 > t, \tau_2 = j))\end{aligned}$$

and

$$\mathrm{PFA}(t) = P(t_{\mathrm{A}} = t | \tau_1 > t, \tau_2 > t) P(\tau_1 > t, \tau_2 > t).$$

For independent geometrically distributed change processes with the same intensity ν, the alarm probabilities simplify. If also the distributions between which the changes appear are the same for the two variables as in the example, we get

$$\begin{aligned}\mathrm{PMA}(t) = &\sum_{i=1}^{t} \sum_{j=1}^{t} (P(t_{\mathrm{A}} = t | \tau_1 = i, \tau_2 = j) \nu^2 (1 - \nu)^{i+j-2}) \\ &+ 2 \sum_{i=1}^{t} (P(t_{\mathrm{A}} = t | \tau_1 = i, \tau_2 > t) \nu (1 - \nu)^{i+t-1})\end{aligned}$$

and

$$\mathrm{PFA}(t) = P(t_{\mathrm{A}} = t \,| \tau_1 > t, \tau_2 > t\,)(1 - \nu)^{2t}.$$

In Figure 4, the predictive value is illustrated for the parallel and reduction approach, both using EWMA. We can see that the parallel approach has a better PV than the reduction approach. This can be expected since the change points are seldom simultaneous when we have independent processes with low intensities.

For simultaneous changes with $\tau_1 = \tau_2 = \tau$ and τ geometrically distributed with $\nu = 0.01$, we have that the probabilities of a motivated and a false alarm, respectively, are

$$\mathrm{PMA} = \sum_{i=1}^{t} (P(t_{\mathrm{A}} = t | \tau = i) \nu (1 - \nu)^{i-1})$$

and

$$\mathrm{PFA}(t) = P(t_{\mathrm{A}} = t \,| \tau > t\,)(1 - \nu)^{t}.$$

As seen in Figure 5, the reduction approach has a better predictive value than the parallel approach when both processes change at the same time. By comparing Figures 4 and 5, we see

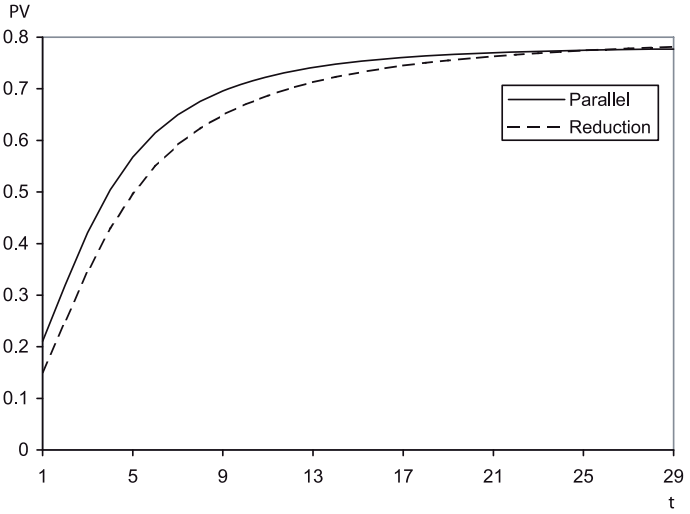

Figure 4. The predictive value (for $C(t) = \{\tau_{\min} \leq t\}$) at different alarm times t_A, for the case where τ_1 and τ_2 are independently geometrically distributed with parameter values $\nu_1 = \nu_2 = 0.01$.

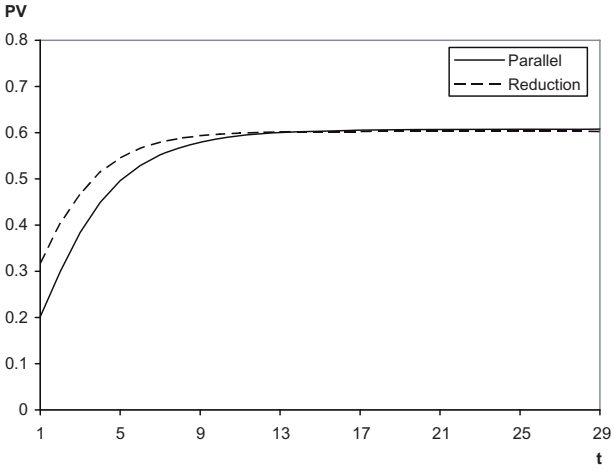

Figure 5. The predictive value at different alarm times t_A, for the case where $\tau_1 = \tau_2 = \tau$ and τ geometrically distributed with $\nu = 0.01$.

that the method that has the best predictive value and thus the most trustworthy alarms depends on the relation between the change points.

5. Discussion

Optimality is often hard to define in multivariate problems due to the several dimensions resulting from the variables. A method could work well for detecting a change in one direction but not in others. In surveillance (univariate as well as multivariate), evaluation is difficult due to the complex time relations. Some methods work well for detecting gradual long-term changes and others for detecting sudden large ones. Thus, it is a challenge to evaluate multivariate surveillance methods

that involve difficulties with both several dimensions and complex time relations. The use of multivariate surveillance methods is growing, and the evaluation challenge has to be approached.

Some new measures, which are generalizations of univariate counterparts, were suggested here and the properties of several measures were demonstrated by applications to various situations. The relation between the change times is very important for deciding which method is the best. For example, the reduction approach gives the shortest delay and the highest predictive value when all processes change at the same time but not when the changes occur separately. The parallel approach has a higher predictive value when the changes are not prone to occur simultaneously. The optimality of the reduction approach for simultaneous changes is a general result and here it is illustrated for a simple example.

It was demonstrated that zero-state and steady-state ARL, which are widely used in univariate surveillance, should be used with care in multivariate surveillance. Unfortunately, the more elaborated CED measure is necessary for full information. This general statement is supported by the example where it is demonstrated that the new suggested measures will reveal important differences between situations which will not be revealed by the conventional measures.

The numerical values of the evaluation measures can be hard to obtain analytically for surveillance methods. Thus, Monte Carlo simulations (as in this paper and many others) or numerical approximations [18] are useful. Evaluation by application to a single case might be interesting but has the drawback of being highly dependent on stochastic variation. Applications to several cases diminish this drawback. An approach between the application to a single case and simulations is the technique of using an observed data series as a start and inducing simulated disturbances to this series [24].

For the measures PFA, ED, and PV, we need the distribution of $\tau_{\min}$, which in turn depends on the distributions of the change times for all processes. These measures are only suitable when the change process is considered to be stochastic. The other measures are also suitable when the change points are considered as unknown but fixed values.

Even if it is appropriate for the application to consider the change points as stochastic, the exact distribution is seldom known. However, any indication about the predictive value is of great importance for the interpretation of an alarm. An alarm does not give cause for extensive action if the predictive value is low. In Figure 4, we can see that the predictive value can be low for early alarms. This means that these should not call for the same actions as later alarms.

Acknowledgements

Kjell Pettersson has given constructive comments throughout the work. The work was supported by the Swedish Emergency Management Agency (grant 0314/206). The authors have declared no conflict of interest.

References

[1] E. Andersson, *Monitoring cyclical processes: A nonparametric approach*, J. Appl. Statist. 29 (2002), pp. 973–990.
[2] E. Andersson, *Effect of dependency in systems for multivariate surveillance*, Comm. Statist. Simulation Comput. 38 (2009), pp. 454–472.
[3] D.W. Apley and F. Tsung, *The autoregressive T-squared chart for monitoring univariate autocorrelated processes*, J. Qual. Technol. 34 (2002), pp. 80–96.
[4] M. Basseville and I. Nikiforov, *Detection of Abrupt changes: Theory and Application*, Prentice Hall, Englewood Cliffs, 1993.
[5] Y. Benjamini and Y. Hochberg, *Controlling the false discovery rate: A practical and powerful approach to multiple testing*, J. R. Stat. Soc. Ser. B 57 (1995), pp. 289–300.
[6] S. Bersimis, S. Psarakis, and J. Panaretos, *Multivariate statistical process control charts: An overview*, Qual. Reliab. Eng. Int. 23 (2007), pp. 517–543.
[7] D. Bock, *Aspects on the control of false alarms in statistical surveillance and the impact on the return of financial decision systems*, J. Appl. Statist. 35 (2008), pp. 213–227.

[8] M. Frisén, *Evaluations of methods for statistical surveillance*, Stat. Med. 11 (1992), pp. 1489–1502.
[9] M. Frisén, *Statistical surveillance: Optimality and methods*, Int. Stat. Rev. 71 (2003), pp. 403–434.
[10] M. Frisén, *Principles for multivariate surveillance*, in *Frontiers in Statistical Quality Control*, H.-J. Lenz and P.-T. Wilrich, eds., to appear.
[11] M. Frisén and C. Sonesson, *Optimal surveillance based on exponentially weighted moving averages*, Sequential Anal. 25 (2006), pp. 379–403.
[12] M. Frisén and P. Wessman, *Quality improvement by likelihood ratio methods for surveillance*, in *Quality Improvement Through Statistical Methods*, B. Abraham, ed., Birkhauser, Boston, 1998, pp. 187–193.
[13] V. Golosnoy, W. Schmid, and I. Yatsyshynets, *Sequential monitoring of optimal portfolio weights*, in *Financial Surveillance*, M. Frisén, ed., Wiley, Chichester, 2007, pp. 179–210.
[14] S. Hao, S. Zhou, and Y. Ding, *Multivariate process variability monitoring through projection*, J. Qual. Technol. 40 (2008), pp. 214–226.
[15] D.M. Hawkins, *Multivariate quality control based on regression-adjusted variables*, Technometrics 33 (1991), pp. 61–75.
[16] Y. Hochberg and A.C. Tamhane, *Multiple Comparison Procedures*, Wiley, New York, 1987.
[17] E. Järpe, *On univariate and spatial surveillance*, Ph.D Thesis, Göteborg University, 2000.
[18] S. Knoth, *The art of evaluating monitoring schemes: How to measure the performance of control charts?* in *Frontiers of Statistical Quality Control*, H.-J. Lenz and P.-T. Wilrich, eds., Physica Verlag, Warsaw, 2006, pp. 74–99.
[19] S. Knoth and W. Schmid, *Monitoring the mean and the variance of a stationary process*, Statist. Neerlandica 56 (2002), pp. 77–100.
[20] G. Lorden, *Procedures for reacting to a change in distribution*, Ann. Math. Statist. 42 (1971), pp. 1897–1908.
[21] C.A. Lowry and D.C. Montgomery, *A review of multivariate control charts*, IIE Trans. 27 (1995), pp. 800–810.
[22] C.W. Lu and M.R. Reynolds Jr, *EWMA control charts for monitoring the mean of autocorrelated processes*, J. Qual. Technol. 31 (1999), pp. 166–188.
[23] C. Marshall et al., *Statistical issues in the prospective monitoring of health outcomes across multiple units*, Journal of the Royal Statistical Society A 167 (2004), pp. 541-559.
[24] M. Mohtashemi, K. Kleinman, and W.K. Yih, *Multi-syndrome analysis of time series using PCA: A new concept for outbreak investigation*, Stat. Med. 26 (2007), pp. 5203–5224.
[25] Y. Okhrin and W. Schmid, *Surveillance of univariate and multivariate nonlinear time series*, in *Financial Surveillance*, M. Frisén, ed., Wiley, Chichester, 2007, pp. 153–177.
[26] M.R. Reynolds Jr and K. Kim, *Multivariate control charts for monitoring the process mean and variability using sequential sampling*, Sequential Anal. 26 (2007), pp. 283–315.
[27] S.W. Roberts, *A comparison of some control chart procedures*, Technometrics 8 (1966), pp. 411–430.
[28] H. Rolka et al., *Issues in applied statistics for public health bioterrorism surveillance using multiple data streams: Research needs*, Stat. Med. 26 (2007), pp. 1834–1856.
[29] G.C. Runger, *Projections and the U^2 chart for multivariate statistical process control*, J. Qual. Technol. 28 (1996), pp. 313–319.
[30] G.C. Runger et al., *Optimal monitoring of multivariate data for fault detection*, J. Qual. Technol. 39 (2007), pp. 159–172.
[31] T.P. Ryan, *Statistical Methods for Quality Improvement*, 2nd ed, Wiley, New York, 2000.
[32] N.S. Sahni, A.H. Aastveit, and T. Naes, *In-Line process and product control using spectroscopy and multivariate calibration*, J. Qual. Technol. 37 (2005), pp. 1–20.
[33] W.A. Shewhart, *Economic Control of Quality of Manufactured Product*, MacMillan, London, 1931.
[34] A.N. Shiryaev, *On optimum methods in quickest detection problems*, Theory Probab. Appl. 8 (1963), pp. 22–46.
[35] C. Sonesson, *Evaluations of some exponentially weighted moving average methods*, J. Appl. Statist. 30 (2003), pp. 1115–1133.
[36] C. Sonesson and M. Frisén, *Multivariate surveillance.*, in *Spatial Surveillance for Public Health*, A. Lawson and K. Kleinman, eds., Wiley, New York, 2005, pp. 169–186.
[37] K. Wärmefjord, *Multivariate quality control and diagnosis of sources of variation in assembled products*, Licentiat Thesis, Göteborg University, 2004.
[38] P. Wessman, *Some principles for surveillance adopted for multivariate processes with a common change point*, Comm. Statist. Theory Methods 27 (1998), pp. 1143–1161.
[39] P. Wessman, *The surveillance of several processes with different change points*, 1999:2, Department of Statistics, Gothenburg University, Göteborg, Sweden, 1999.
[40] W.H. Woodall and S. Amiriparian, *On the economic design of multivariate control charts*, Comm. Statist. Theory Methods 31 (2002), pp. 1665–1673.
[41] S. Zacks and R.S. Kenett, *Process tracking of time series with change points*, in *Recent Advances in Statistics and Probability*, J.P. Vilaplana and M.L. Puri, eds., International Science Publishers, Zeist, The Netherlands, 1994, pp. 155–171.

Paper III

To appear in *Communication in Statistics - Theory and Methods*

Sufficient reduction in multivariate surveillance

Marianne Frisén[1], Eva Andersson[1,2], and Linus Schiöler[1]

[1]Statistical Research Unit, University of Gothenburg, Göteborg, Sweden

[2]Occupational and environmental medicine, Sahlgrenska Academy and Sahlgrenska University Hospital, University of Gothenburg, Göteborg, Sweden

The relation between change points in multivariate surveillance is important but seldom considered. The sufficiency principle is here used to clarify the structure of some problems, to find efficient methods, and to determine appropriate evaluation metrics. We study processes where the changes occur simultaneously or with known time lags. The surveillance of spatial data is one example where known time lags can be of interest. A general version of a theorem for the sufficient reduction of processes that change with known time lags is given. A simulation study illustrates the benefits or the methods based on the sufficient statistics.

Keywords change-points, exponential family, MEWMA, monitoring, inference principles

1. Introduction

In society there is a great need for continuous surveillance of processes with the aim of detecting an important change in the underlying process as soon as possible after the change has occurred. The inference is quite different in on-line surveillance as compared to hypothesis testing. In surveillance there are no fixed hypotheses. Even if the situation is stable at the current time, a change can happen later. Timeliness is important in surveillance. Since the probability of a false alarm increases with time and tends to one for most surveillance methods, evaluation by significance level, power, and other well-known metrics is not useful for ordinary surveillance problems. Some surveillance methods have been constructed to resemble methods for hypothesis testing, see for example Chu, Stinchcombe, & White, (1996). These methods are constructed to have a false alarm probability less than one. This could be an advantage, since it allows statements like those in hypothesis testing to be made. However, Frisén, (2003), Aue & Horvath, (2004), and Bock, (2008) demonstrated that the detection ability of these methods declines rapidly for late changes. These methods are suitable only for applications where a possible change appears at or soon after the start. Sometimes methods like the CUSUM method by Page, (1954) or the Shiryaev-Roberts method by Shiryaev,

Address correspondence to Marianne Frisén, Statistical Research Unit, University of Gothenburg, PO Box 640, SE 40530 Gothenburg, Sweden; E-mail: marianne.frisen@statistics.gu.se.

(1963), which were constructed to be optimal for on-line surveillance, are demonstrated to be useful also for retrospective hypothesis testing, as in Lee, Ha, Na, & Na, (2003) and Vexler & Wu, (2009). There are problems situated between hypothesis testing and surveillance, but in this paper we will deal only with inference suitable for on-line surveillance.

The first versions of modern control charts (Shewhart, (1931)) were made for industrial use. Multivariate surveillance is of interest in industrial production, for example in order to monitor the multiple sources of variation in assembled products. Wärmefjord, (2004) described the multivariate problem for the assembly process of the Saab automobile. In recent years, there has been an increased interest in statistical surveillance also in other areas than industrial production. The increased interest in surveillance methodology in the US following the 9/11 terrorist attack is notable. In the US, as well as in other countries, several new types of data are now being collected. Since the collected data involve several related variables, this calls for multivariate surveillance techniques. The surveillance of several parameters of one distribution (such as the mean and the variance of a normal distribution), see for example Knoth & Schmid, (2002), can involve the same problems as the surveillance of a multidimensional distribution originating from the observation of different variables. Spatial surveillance is useful for the detection of a local change or a spread. One example is the spread of a disease such as influenza, as in Schiöler, (2008) and Frisén, Andersson, & Schiöler, (2009b). Another example is the spread of a harmful agent such as nuclear radiation, as in Järpe, (2001). Spatial surveillance is multivariate since several locations are involved. Recently, there have also been efforts to use multivariate surveillance for financial decision strategies (see for example Okhrin & Schmid, (2007) and Golosnoy, Schmid, & Okhrin, (2007)) with respect to various assets.

Reviews on multivariate surveillance methods can be found for example in Basseville & Nikiforov, (1993), Ryan, (2000), Frisén, (2003), Sonesson & Frisén, (2005), Bersimis, Psarakis, & Panaretos, (2007), and Frisén, (2009). Optimality is hard to derive and sometimes even hard to define in multivariate problems. However, we will demonstrate how the structure of some multivariate surveillance problems can be simplified by the sufficiency principle and how this will lead to more efficient methods than those suggested earlier.

At each time point, a new observation is made on the process. The p-variate process under surveillance is denoted by $\mathbf{Y} = \{\mathbf{Y}(t),\ t = 1,2,...\}$, where $\mathbf{Y}(t) = \{Y_1(t),\ Y_2(t),...,\ Y_p(t)\}$. We aim to detect the change from a stable state D to a harmful state C as soon as possible after the change has occurred, in order to give warnings and take corrective actions. At decision time s

we base the decision on the available information $\mathbf{Y}^s = \{\mathbf{Y}(1), \mathbf{Y}(2)... \mathbf{Y}(s)\}$ and use the observation vector $\mathbf{Y}^s$ to form an alarm statistic. An alarm is called the first time that the statistic exceeds an alarm limit. In the univariate case, the change happens at the unknown time point τ. In both theoretical works and applications the change point, τ, is sometimes considered as an unknown constant and other times as a stochastic variable. Both views have their merits, thus we will first give the results for τ considered as fixed and then demonstrate that the Theorem holds also for stochastic τ. In the multivariate case, we observe p processes which can change at different times τ_1, .. τ_p. Here an important aim is to detect the first time that not all processes are in control, that is, we want to make inference about $\tau_{\min} = \min\{\tau_1,...,\tau_p\}$. If no change ever occurs in process i, we denote this by "$\tau_i = \infty$". We consider models where the observations $Y_i(t)$ and $Y_i(t+j)$ are independent, given the values of the change points, and for each variable, i, there is one distribution, with density $f_i^0(t)$, for $t < \tau_i$ and another, with density $f_i^1(t)$, for $t \geq \tau_i$. In this paper we concentrate on the one-parameter exponential family.

In Section 2 different approaches to the construction of multivariate surveillance methods are described and exemplified. Theoretical results on sufficient reduction are given in Section 3. In Section 4 we discuss the challenges of evaluating multivariate surveillance methods with special focus on how the structure of the multivariate problem is clarified by the sufficiency principle. In Section 5 we illustrate the theory by a simulation study. Concluding remarks are made in Section 6.

2. Approaches to multivariate surveillance

Some commonly used general approaches for adapting univariate methods to multivariate surveillance will be described and exemplified. Principal differences between approaches for handling multivariate data in surveillance will be demonstrated.

2.1. *Dimension reduction*

In Statistical Process Control (SPC) it is practical to use only one control chart instead of several. Thus many suggestions have been made on reduction to one chart (see e.g. Cheng & Thaga, (2006)). A stepwise reduction of the multivariate surveillance problem is natural. An easy way to simplify the situation is to reduce the p-variate vector at each time point into one

statistic and then use a system for univariate surveillance on this statistic. One example is the suggestion by Crosier, (1988) to summarize data by the Hotelling T variable and then apply the univariate CUSUM method to the T variable, making it a scalar accumulation method. As we will describe in Section 3, a sufficient dimension reduction can be found for some situations.

2.2. *Parallel surveillance*

A stepwise solution of the multivariate surveillance problem can alternatively be accomplished by monitoring each variable separately. The approach with parallel systems is often called "combined univariate" methods or "parallel" methods. The most common way to combine the information from several univariate methods is to signal an alarm at the first time that any of the univariate methods gives an alarm. This is a special case of the union-intersection technique suggested by Roy, (1953).

2.3. *Vector accumulation*

The accumulated information on each component is utilized by a transformation of the vector of component-wise alarm statistics into a scalar alarm statistic. Thus a surveillance method is applied to each of the p processes, resulting in p-variate alarm statistics at each decision time s. This p-variate statistic is then transformed into a scalar, which is the alarm statistic for the whole system at time s. An alarm is triggered if this statistic exceeds a limit. As an example, Lowry, Woodall, Champ, & Rigdon, (1992) proposed a multivariate extension, MEWMA, of the univariate EWMA. The MEWMA method uses a vector of univariate EWMA statistics. For each variable Y_j and each time t, we have the EWMA statistic $Z(t)=\lambda_j Y_j(t)+(1-\lambda_j)Z_j(t-1)$ where $Z(0)=0$. At decision time s we have $\mathbf{Z}(s)=\{Z(\mathbf{y}_1^s), Z(\mathbf{y}_2^s),..., Z(\mathbf{y}_p^s)\}$. An alarm is triggered at $t_A = \min\{s;\quad \mathbf{Z}(s)^T \boldsymbol{\Sigma}_{\mathbf{Z}(s)}^{-1} \mathbf{Z}(s) > L\}$. The properties of the method are described in Section 5.2.2. Vector accumulation methods based on CUSUM have also been proposed, but there are several possibilities of how to handle the characteristic barrier of the CUSUM methods (see Sonesson & Frisén, (2005)).

2.4. *Joint solution*

A joint solution of the original full problem, without stepwise solutions, is preferred when possible. In general, the set of likelihood ratios $f_1(y)/f_0(y),...,f_q(y)/f_0(y)$ is sufficient for the problem (see for example Cox & Hinkley, (1974, p. 21)). It follows that the set of partial likelihood ratios is sufficient for surveillance problems:

$$L(s,m_1,...m_p)=\frac{f(\mathbf{Y}_1^s,...\mathbf{Y}_p^s\,|\tau_1=m_1,...\tau_p=m_p)}{f(\mathbf{Y}_1^s,...\mathbf{Y}_p^s\,|\tau_1>s,...\tau_p>s)}\ .$$

The full likelihood ratio method for the multivariate problem (see for example Andersson, (2009)) requires knowledge of the distribution of the change times. When the full likelihood for $\mathbf{Y}^s = \{\mathbf{Y}(1), \mathbf{Y}(2)... \mathbf{Y}(s)\}$ is available, it provides a good basis for surveillance since optimal methods are mostly constructed based on the likelihood. However, the full likelihood can be complicated for some problems, and therefore a reduction may be considered. A sufficient reduction will not reduce the information, but other reductions will. A jointly optimal solution can be constructed by a sufficient reduction (where no information is lost in the reduction step), followed by an optimal surveillance method applied to the reduced statistic. Stepwise approaches which start with a reduction (either in time or in the variables) and then use a possibly optimal univariate method can be suspected to be suboptimal. Only reductions which are sufficient can be expected to result in jointly optimal solutions, since no information is lost.

3. Sufficient reduction

A statistic T is sufficient for a family of distributions if and only if $f_{Y|T}(y|t)$ is the same for all distributions belonging to the family $\mathcal{F}$ of interest (see for example Cox & Hinkley, (1974)). A sequence $T^1(Y_1)$, $T^2(Y_2)$,… is a sufficient sequence of statistics for the distributional families $\mathcal{F}^1$, $\mathcal{F}^2$ **…** if for all s**,** $T^s(Y_s)$ is a sufficient statistic for the family $\mathcal{F}^s$.

For a shift at τ in a univariate distribution between two fully specified distributions, the set of likelihood ratios $L(s,t) = f_Y^s(\mathbf{Y}^s\,|\tau=t)\,/f_Y^s(\mathbf{Y}^s\,|D)$ is sufficient for the distributional family of $\mathbf{Y}^s$ defined by the time of change τ.

According to the sufficiency principle, all conclusions to be drawn should depend on one sufficient statistic only.

3.1. *Simultaneous changes*

Consider the case where all processes have the same change point so that $\tau_1 = \tau_2 = \ldots \tau_{p=} \tau$. An example could be when all variables are indicators of the same phenomena. In most evaluations of multivariate surveillance it is assumed that all changes are simultaneous. For a change at τ between the distributions f^0 and f^1 we have the distribution for the s observations

$$f(\mathbf{Y}^s \mid \tau = m) = \prod_{t=1}^{m-1} f^0(Y(t)) \prod_{t=m}^{s} f^1(Y(t)) = \prod_{t=1}^{s} f^0(Y(t)) \prod_{t=m}^{s} \frac{f^1(Y(t))}{f^0(Y(t))} .$$

It now becomes possible to identify the separate factors: the part that depends on the data (but not the value of τ) as well as the part that depends on the s-dimentional vector $L^s(\mathbf{Y}^s)=\{L(s,m), t=1,...s\}$, where m is the common change time. Thus $L^s(\mathbf{Y}^s)$ is sufficient for the distributional family for each s. From this it follows that the sequence of s likelihood ratios is a sufficient sequence. This was proven by Wessman, (1998) both for a fixed unknown value of τ and for a stochastic time of change. When the aim is to detect a fully specified, simultaneous change in a multivariate process and the distributions before and after the change are fully specified, it is possible to construct a univariate surveillance procedure based on the sufficient sequence of likelihood ratios. Examples will be given in the next section as special cases of the general theorem in the next section.

3.2. *Changes with time lags*

We will now consider the case where there are known time lags between the changes of the p processes. We also discuss in which type of problems a known time lag and some knowledge on the size of this lag may be available. See Section 6 for a discussion on the effect of imperfect knowledge on the time lag. For each application the choice of distributional models must be carefully considered.

There may in some cases be one source of information of good quality that is available after a delay and another source with worse quality that is available early. The multivariate utilization of these data sets might benefit from information on how large the time lag is.

Another example is the spatial spread of a disease. In Schiöler, (2008) analyses are made of Swedish influenza data and it is shown that the influenza spreads from the larger cities (the Metropolitan areas) to the rest of the country (Local) with a lag of approximately 2 weeks.

A third example of monitoring of a succession of events is a surveillance system for radioactivity emerging from a nuclear plant. The radioactivity reaches the measuring locations with a time lag which is proportional to the distance from the source. Järpe, (2000) studied measurements at different geographical locations in Sweden. Several models for the spread of radioactive material by the wind were studied. At each location, the radioactivity increased with a time lag which was assumed to be proportional to the distance from the nuclear plant. For the situation with a shift of equal size in the expected value of Gaussian processes, when the shifts occur with known lags and where we have independent (given the change points) normal distributions with the same variances, Järpe, (2000) demonstrated that a sufficient reduction to univariate surveillance exists. Here we will prove that a sufficient reduction to a univariate statistic exists as long as the processes belong to the one-parameter exponential family.

Theorem

Consider p processes Y_1, Y_2, ..., Y_p which all belong to the one-parameter exponential family and which are independent conditional on the change points (independent over time as well as across coordinates). We consider both the situation when the time of change is fixed but unknown and also a stochastic time of change. There exists a sufficient reduction of the set of observation vectors $\{\mathbf{y}_1, \mathbf{y}_2, \ldots \mathbf{y}_p\}$ to a univariate statistic for the detection of shifts in the parameter vector when the changes occur with known time lags ($q_2, q_3, \ldots, q_p$) where $q_i=\tau_i-\tau_{i-1}$. A sufficient statistic for the detection of shifts of sizes $\delta_1, \delta_2, \ldots \delta_s$ is the set

$\delta_1 y_1(t) + \delta_2 y_2(t+q_2) + \ldots + \delta_p y_p(t+q_p)$, for $1 \le t \le s-q_2-q_3-\ldots-q_p$

$\delta_1 y_1(t) + \delta_2 y_2(t+q_2) + \ldots + \delta_{p-1} y_{p-1}(t+q_{p-1})$, for $s-q_2-q_3-\ldots-q_p < t \le s-q_2-q_3-\ldots-q_{p-2}$,

…

$\delta_1 y_1(t) + \delta_2 y_2(t+q_2)$, for $s-q_2-q_3 < t \le s-q_2$,

$y_1(t)$, for $s-q_2 < t \le s$.

Proof

Since the observations are independent given the values of the change points, the distribution can be written as a product. We will first consider a fixed unknown value of $\tau_{\min.}$ The likelihood expressions for the exponential family can be written as

$$
\begin{aligned}
&f(Y|\tau_{\min} \le s) = \\
&\exp\left\{\sum_{t=1}^{\tau_1-1}\sum_{j=1}^{p}\left[y_j(t)\varphi_j + g(\varphi_j) + h(y_j(t))\right]\right\}\cdot \\
&\exp\left\{\sum_{t=\tau_1}^{\tau_2-1}\left[\sum_{j=1}^{1}\left[y_j(t)(\varphi_j+\delta_j) + g(\varphi_j+\delta_j) + h(y_j(t))\right] + \sum_{j=2}^{p}\left[y_j(t)\varphi_j + g(\varphi_j) + h(y_j(t))\right]\right]\right\}\cdot \\
&\cdot\ldots\cdot\exp\left\{\sum_{t=\tau_{p-1}}^{\tau_p-1}\left[\sum_{j=1}^{p-1}\left[y_j(t)(\varphi_j+\delta_j) + g(\varphi_j+\delta_j) + h(y_j(t))\right] + \sum_{j=p}^{p}\left[y_j(t)\varphi_j + g(\varphi_j) + h(y_j(t))\right]\right]\right\}\cdot \\
&\exp\left\{\sum_{t=\tau_p}^{s}\sum_{j=1}^{p}\left[y_j(t)(\varphi_j+\delta_j) + g(\varphi_j+\delta_j) + h(y_j(t))\right]\right\}
\end{aligned}
$$

and

$$
\begin{aligned}
&f(Y|\tau_{\min} > s) = \\
&\exp\left\{\sum_{t=1}^{s}\sum_{j=1}^{p}\left[y_j(t)\varphi_j + g(\varphi_j) + h(y_j(t))\right]\right\}
\end{aligned}
$$

The likelihood ratio, conditional on $\tau_{\min}$=m, equals $L(s,m) = \dfrac{f(Y \mid \tau_{\min} = m \le s)}{f(Y \mid \tau_{\min} > s)}$ and thus the log likelihood ratio is

$$\sum_{t=1}^{m-1}\sum_{j=1}^{p}\left[y_j(t)\varphi_j + g(\varphi_j) + h(y_j(t))\right] +$$
$$\sum_{t=m}^{m+q_2-1}\left[\sum_{j=1}^{1}\left[y_j(t)(\varphi_j+\delta_j) + g(\varphi_j+\delta_j) + h(y_j(t))\right] + \sum_{j=2}^{p}\left[y_j(t)\varphi_j + g(\varphi_j) + h(y_j(t))\right]\right] +$$
$$\sum_{t=m+q_2}^{m+q_2+q_3-1}\left[\sum_{j=1}^{2}\left[y_j(t)(\varphi_j+\delta_j) + g(\varphi_j+\delta_j) + h(y_j(t))\right] + \sum_{j=3}^{p}\left[y_j(t)\varphi_j + g(\varphi_j) + h(y_j(t))\right]\right] +$$
$$+\ldots+$$
$$+\sum_{t=m+q_2+\ldots+q_{p-1}}^{m+q_2+\ldots+q_p-1}\left[\sum_{j=1}^{p-1}\left[y_j(t)(\varphi_j+\delta_j) + g(\varphi_j+\delta_j) + h(y_j(t))\right] + \sum_{j=p}^{p}\left[y_j(t)\varphi_j + g(\varphi_j) + h(y_j(t))\right]\right] +$$
$$+\sum_{t=m+q_2+\ldots+q_p}^{s}\sum_{j=1}^{p}\left[y_j(t)(\varphi_j+\delta_j) + g(\varphi_j+\delta_j) + h(y_j(t))\right] -$$
$$-\sum_{t=1}^{s}\sum_{j=1}^{p}\left[y_j(t)\varphi_j + g(\varphi_j) + h(y_j(t))\right]$$

This can be arranged into

$$\sum_{t=m}^{s-q_2-\ldots-q_p} y_1(t)\delta_1 + \sum_{t=s-q_2-\ldots-q_p+1}^{s-q_2-\ldots-q_{p-1}} y_1(t)\delta_1 + \ldots + \sum_{t=s-q_2-q_3+1}^{s-q_2} y_1(t)\delta_1 + \sum_{t=s-q_2+1}^{s} y_1(t)\delta_1$$
$$+\sum_{t=m+q_2}^{s-q_3-\ldots-q_p} y_2(t)\delta_2 + \sum_{t=s-q_3-\ldots-q_p+1}^{s-q_3-\ldots-q_{p-1}} y_2(t)\delta_2 + \ldots + \sum_{t=s-q_3+1}^{s} y_2(t)\delta_2$$
$$+\ldots+\sum_{t=m+q_2+\ldots+q_{p-1}}^{s-q_p} y_{p-1}(t)\delta_{p-1} + \sum_{t=s-q_p+1}^{s} y_{p-1}(t)\delta_{p-1}$$
$$+\sum_{t=m+q_2+\ldots+q_p}^{s} y_p(t)\delta_p$$
$$+z(\delta_1,\delta_2,\ldots,\delta_p,\varphi_1,\varphi_2,\ldots,\varphi_p)$$

where $z(\delta_1, \delta_2,\ldots \delta_p, \varphi_1, \varphi_2,\ldots \varphi_p) =$

$$\sum_{t=m}^{s}\left(g(\varphi_1+\delta_1)-g(\varphi_1)\right)$$
$$+\sum_{t=m+q_2}^{s}\left(g(\varphi_2+\delta_2)-g(\varphi_2)\right)$$
$$+\ldots+\sum_{t=m+q_2+\ldots+q_{p-1}}^{s}\left(g(\varphi_{p-1}+\delta_{p-1})-g(\varphi_{p-1})\right)$$
$$+\sum_{t=m+q_2+\ldots+q_p}^{s}\left(g(\varphi_p+\delta_p)-g(\varphi_p)\right)$$

is independent of the observations. The expression above can be rewritten as

$$\sum_{t=m}^{s-q_2-\ldots-q_p}\left[y_1(t)\delta_1+\left(y_2(t+q_2)\delta_2\right)+\ldots+y_{p-1}(t+q_2+\ldots+q_{p-1})\delta_{p-1}+y_p(t+q_2+\ldots+q_p)\delta_p\right]$$
$$+\sum_{t=s-q_2-\ldots-q_p+1}^{s-q_2-\ldots-q_{p-1}}\left[y_1(t)\delta_1+y_2(t+q_2)\delta_1+\ldots+y_{p-1}(t+q_2+\ldots+q_{p-1})\delta_{p-1}\right]$$
$$+\ldots+\sum_{t=s-q_2-q_3+1}^{s-q_2}\left[y_1(t)\delta_1+y_2(t+q_2)\delta_2\right]$$
$$+\sum_{t=s-q_2+1}^{s}y_1(t)\delta_1$$
$$+z(\delta_1,\delta_2,\ldots,\delta_p,\varphi_1,\varphi_2,\ldots,\varphi_p)$$

Thus logL(s,m) is a one-one function of the statistic in the Theorem, and thus it is a sufficient statistic for L(s,m) and thus for the problem.

If τ_{min} is stochastic with some density g(t), then the density of Y^s can be written:

$$f(\mathbf{Y})=\sum_{t=1}^{\infty}g(t)f(\mathbf{Y}\mid\tau=t)\,.$$

This is a simple function of $f(\mathbf{Y}\mid\tau=t)$ and hence the arguments above can be used to show that the statistic in the Theorem is sufficient for the problem also for this case.

The Theorem is general and thus has many parameters. In order to illustrate the idea we will now look at some special cases. The performance for these special cases will be illustrated in Section 5.

Corollary 1

A special case of the Theorem concerns two processes (p=2) when the changes occur at the same time (q=0). In this situation we have by the Theorem that $\delta_1 Y_1(t) + \delta_2 Y_2(t)$ for $1 \leq t \leq s$ is sufficient. If, for example, $\delta_1 = 2\delta_2$ we have that $2\delta_2 Y_1(t) + \delta_2 Y_2(t)$ is sufficient. From this it follows that the statistic $\frac{2}{3} Y_1(t) + \frac{1}{3} Y_2(t)$ is sufficient. If we have equal shifts in the parameter vector ($\delta_1 = \delta_2 = \delta$), then $\delta Y_1(t) + \delta Y_2(t)$ is a sufficient statistic. From this it follows that the set of means of the observations

$$\text{SuffR}^0(t) = \frac{Y_1(t) + Y_2(t)}{2}$$

is sufficient.

Corollary 2

Another special case of the Theorem concerns two processes (p=2) which have equal shifts in the parameter vector ($\delta_1 = \delta_2 = \delta$) and where the changes occur with a known time lag q. In this situation we have, by the Theorem, that a sufficient statistic is the set

$\delta\left(Y_1(t) + Y_2(t+q)\right)$ for t=1,…s-q,

$\delta Y_1(t)$ for t=s-q+1, …s

We need two arguments to specify the statistic when q>0, since the series changes when s increases. For q=1 a sufficient statistic is the set

$\{\text{SuffR}^1(s, t)\}$, for t=1, 2, … s.

Thus, for s=1, the sufficient set is

$\{\text{SuffR}^1(1, 1) = Y_1(1)\}$.

For s=2, the sufficient set is

$$\{\text{SuffR}^1(2,1)=\left(Y_1(1)+Y_2(2)\right)/2,\ \text{SuffR}^1(2,2)=Y_1(2)\ \}.$$

For s=3, the sufficient set is

$$\{\text{SuffR}^1(3,1)=\left(Y_1(1)+Y_2(2)\right)/2,\ \text{SuffR}^1(3,2)=\left(Y_1(2)+Y_2(3)\right)/2,\ \text{SuffR}^1(3,3)=Y_1(3)\ \}.$$

For q=5, a sufficient statistic is the set

$$\{\text{SuffR}^5(s,t)=\ \{\left(Y_1(t)+Y_2(t+5)\right)/2,\ldots\left(Y_1(s-5)+Y_2(s)\right)/2,\quad Y_1(s-4),\ldots Y_1(s)\}\,,$$

for t=1, 2…s.

The main theory of statistical surveillance is constructed for a change between two distributions – one for $t<\tau_i$ and another for $t\geq\tau_i$. The $\text{SuffR}^q(s,t)$ statistic does not necessarily change between two distributions for $q>0$. For iid Gaussian distributions (conditional on τ_i) with expected values μ^0 for $t<\tau_i$ and μ^1 for $t\geq\tau_i$, and constant variance σ^2, the distributions of the sufficient $\text{SuffR}^q(s,t)$ statistics have the expected value μ^0 for $t< \tau_{min}$ and μ^1 for $t\geq \tau_{min}$. However, the variance is not the same for $t>q$ as for $t\leq q$. For example, for a lag of 1, the variance for $\text{SuffR}^1(2,1)$ equals $\sigma^2/2$, whereas the variance for $\text{SuffR}^1(2,2)$ equals σ^2. Other transformations, which are also sufficient, could be considered. One alternative is to divide the sums in the sufficient statistic SuffR^q with $\sqrt{2}$ instead of 2. This results in a constant variance for all components but not constant expected values. For $t\geq \tau_{min}$ the expected value shifts from $\sqrt{2}\mu^1$ for the first components of the series to μ^1 (for the last components). This seems like a larger drawback, and we will thus study the $\text{SuffR}^q(s,t)$ statistic in the examples in Section 5.4. In spite of the fact that we cannot rely on theoretical optimality (since the SuffR^q statistic does change between more than two distributions), we will see that the statistic works well.

4. Evaluation

4.1. *Optimality*

It can be difficult to find a definition of optimality that holds for all different aspects of multivariate problems in surveillance, see Frisén, (2003). In multivariate problems there are always many dimensions to consider. In surveillance there is the additional complexity of the different relations between the change points, ranging from simultaneous changes to independent changes. Nevertheless, sufficient reductions make it possible to find optimal solutions for at least one important situation.

After sufficient reduction to a univariate statistic, we can use earlier optimality results of univariate surveillance. Different combinations of the partial likelihood ratios are known to have different optimality properties, as described by Frisén, (2003). In Frisén & de Maré, (1991) it is shown that the full likelihood ratio method, which is a weighted sum of L(s,t), with the weights proportional to $P(\tau=t)$, yields a minimal expected delay in univariate surveillance. This follows from the results by Shiryaev, (1963), where optimality is shown when the change point follows a geometric distribution. Another function of the partial likelihood ratios is the maximum likelihood ratio component L(s,t) with respect to t. This alarm statistic is mini-max optimal, as proved by Moustakides, (1986). The EWMA method was demonstrated by Frisén, (2003) and Frisén & Sonesson, (2006) to be an approximation of the full likelihood ratio method. Here we will use $\lambda=0.35$ as a reasonable value for all methods to make them more comparable but without any claim of optimality.

For simultaneous changes, it was demonstrated in Section 3 that the multivariate problem can be reduced to a univariate problem of a change between two distributions: one for $t<\tau$ and another for $t\geq\tau$. Thus, the ordinary theory of optimal surveillance can be applied. Surveillance of the sufficient statistic by an optimal univariate method is thus optimal for the multivariate problem.

In the multivariate setting with different change points, the full likelihood ratio equals the joint solution. We may be able to find the full likelihood ratio, weighted by the geometric distribution of τ, which in the univariate case guarantees a minimal delay. Sun & Basu, (1995) studied multivariate surveillance with p=2 and used the assumption that (τ_1,τ_2) follows a bivariate geometric distribution. This means that also τ_{min} follows a geometric distribution. If

τ_{min} is considered as the change point, then the requirement of a geometric distribution is satisfied. However, in proofs for optimality such as those of Shiryaev, (1963) and Moustakides, (1986), it is also required that **Y**(t) is independently and identically distributed before as well as after the change point. The requirement of identical distributions is not satisfied for **Y**(t) for all t after τ_{min}, for the situation when there are several change points. Nevertheless, the different types of combinations of partial likelihood expressions (as described above) can be assumed to be suitable for different types of (approximate) optimality. In Section 5, examples will be used to demonstrate that the methods based on the sufficient statistic work well also for situations where optimality cannot be proven.

4.2. *Evaluation measures in multivariate surveillance*

The most commonly used measure of delay of the time, t_A, of the alarm is $ARL^1 = E[t_A \,|\, \tau = 1]$ which is also called the zero state ARL since it is a measure of the delay when the change happens immediately. A measure for the opposite situation, when the change time tends to infinity, is the steady state ARL (see for example Lu & Reynolds Jr, (1999) and Reynolds & Kim, (2007)). In univariate surveillance this measure is unique for specified distributions and a specified method. In a multivariate setting, however, this measure is not unique but depends on the relation between the change points when they tend to infinity. It is common to calculate the measure for the situation of simultaneous changes even if the assumption of simultaneous changes is only implicit. However, as was pointed out in Section 3.1, the situation with simultaneous changes is not a genuine multivariate problem since it can be reduced to a univariate one. As was seen in Section 3.1, there are optimal methods for this situation.

In Frisén, Andersson, & Schiöler, (2009a) the conditional expected delay was recommended for situations with different relations between the τ-values

$$\mathrm{CED}(\tau_1, \tau_2 \ldots \tau_p) = E[t_A - \tau_{\min} \,|\, t_A \geq \tau_{\min}].$$

This measure will be used in the next section to evaluate methods for different situations. The CED is rather constant in the examples below. However, the effect of early or late changes would be more pronounced for a smaller value of λ (and less pronounced for larger values of λ).

5. Examples

In order to illustrate the performance of different multivariate methods, especially those based on reduction, we apply them to a number of different situations. We will concentrate on the way in which the relations between the change times, τ_1, τ_2, ..., τ_p, influence the properties of different surveillance methods. In Section 5.1 we give a simple model which will be used in the simulation study, in Section 5.2 we describe the methods which are compared, and in Sections 5.3 and 5.4, respectively, we report the results for simultaneous changes and changes with different change points.

5.1. *Simple model*

A very simple example with two processes will be used. The two processes, Y_1 and Y_2, are assumed to be independent (conditional on the change times)

$$Y_1(t) \sim \begin{cases} N(0,1) & t < \tau_1 \\ N(2,1) & t \geq \tau_1 \end{cases}$$

$$Y_2(t) \sim \begin{cases} N(0,1) & t < \tau_2 \\ N(2,1) & t \geq \tau_2 \end{cases}$$

5.2. *Methods*

In Section 2 we described how univariate techniques can be generalized to handle multivariate situations. We have chosen the EWMA method as the method for accumulating the information over time, since it is commonly used also in multivariate situations. The EWMA method was introduced in the quality control literature by Roberts, (1959) and has received much attention. As regards the variance of the EWMA statistic there are two versions: the exact and the asymptotic variance. We will use the asymptotic variance, both for simplicity and on the basis of the arguments given in Frisén & Sonesson, (2006) concerning properties. At time s the statistic of the EWMA method for the univariate surveillance of Y is

$$Z(s)=\lambda(1-\lambda)^s \sum_{t=1}^{s} (1-\lambda)^{-t} Y(t),$$

where $0<\lambda\leq 1$ and Z_0 is the target value, which is zero in the examples. The EWMA statistic is a weighted sum of all observations available at the decision time s. Here we choose the value $\lambda=0.35$. For the comparisons we set alarm limits to ensure the same median run length to a false alarm ($MRL^0=100$). We will compare the results of several approaches to multivariate surveillance: i) the EWMA method applied to a sufficient reduction of data, ii) the MEWMA method, iii) a system based on two parallel EMWA methods, and iv) the EWMA method applied to the univariate process that changes first. These methods will now be described.

5.2.1 EWMA based on reduction

If the two processes in Section 5.1 have simultaneous change points ($\tau_1=\tau_2$), then the reduction to the statistic $SuffR^0(t)=(Y_1(t)+Y_2(t))/2$ is sufficient. The EWMA method can then be applied to this statistic. This reduction method is labeled $SuffR^0$ in the figures.

We will also study the reduction $SuffR^5(s,t)$ for the case of a lag of 5 ($\tau_2=\tau_1+5$). In the surveillance process the EWMA is applied to the sufficient statistics, and the time of alarm for the reduction methods is the first time when the EWMA statistic exceeds a constant alarm limit. Note that the recursive formula $Z(s) = (1-\lambda)Z(s-1)+\lambda Y(s)$, for s=1, 2,... , which can be used for a univariate statistic Y, is not always valid here. The whole $SuffR^q(t)$ series is revised at each decision time (except for q=0). Thus the original EWMA $Z(s) = \lambda(1-\lambda)^s \sum_{t=1}^{s}(1-\lambda)^{-t} SuffR^q(t)$ should be used. For lag 5 we have

$$Z(s)=(1-\lambda)^s Z_0+(1-\lambda)^{s-1}\lambda(Y_1(1)+Y_2(6))/2+ (1-\lambda)^{s-2}\lambda(Y_1(2)+Y_2(7))/2 +...$$

$$+ (1-\lambda)^2\lambda(Y_1(s-5)+Y_2(s))/2 +...+ (1-\lambda)^1\lambda Y_1(s-1) + \lambda Y_1(s).$$

5.2.2 MEWMA

MEWMA can be described as a Hotelling T^2 control chart applied to univariate EWMA statistics instead of to the original data and is thus a vector accumulation method. For our simple example and with the value of λ equal for both processes it is

$$EWMA(s) = \frac{Z_1(s)^2 + Z_2(s)^2}{\lambda/(2-\lambda)}.$$

5.2.3 Parallel EWMA

The parallel approach means that the EWMA method is applied to $Y_1(t)$ and $Y_2(t)$ separately. The time of alarm for the Parallel method is the first of either of the alarm times.

5.2.4 Univariate

For comparison we also have the results from the EWMA method applied to only one process. This corresponds to the situation when there is prior knowledge about which process will change first and therefore efficient to monitor only this one. This method is labeled "Univariate" in the diagrams.

5.3. *Results for simultaneous changes*

Below we present the results of the delay curve for the methods described above and the model in Section 5.1. First we study the situation when $\tau_1=\tau_2=\tau_{min}$, for $\tau_{min}=1, 2, .., 15$. By Corollary 1, a method based on the sufficient reduction to the $SuffR^0$ statistic should be used. We compare the EWMA method based on $SuffR^0$ with the MEWMA method and the Parallel method.

In Figure 1 we see that for simultaneous changes, the EWMA method based on reduction to the statistic $SuffR^0(t)= (Y_1(t)+Y_2(t))/2$ gives the shortest delay. This is in accordance with theory, as described in Section 3.1. It may be surprising that the popular MEWMA method gives the worst result. In this simple example, however, the flexibility of the MEWMA method does not constitute an advantage. When using the other methods it is advantageous to know the direction of the change. By contrast, the MEWMA method based on Hotelling T^2 is directionally invariant. There are suggestions of one-sided versions of MEWMA, but they were not used here.

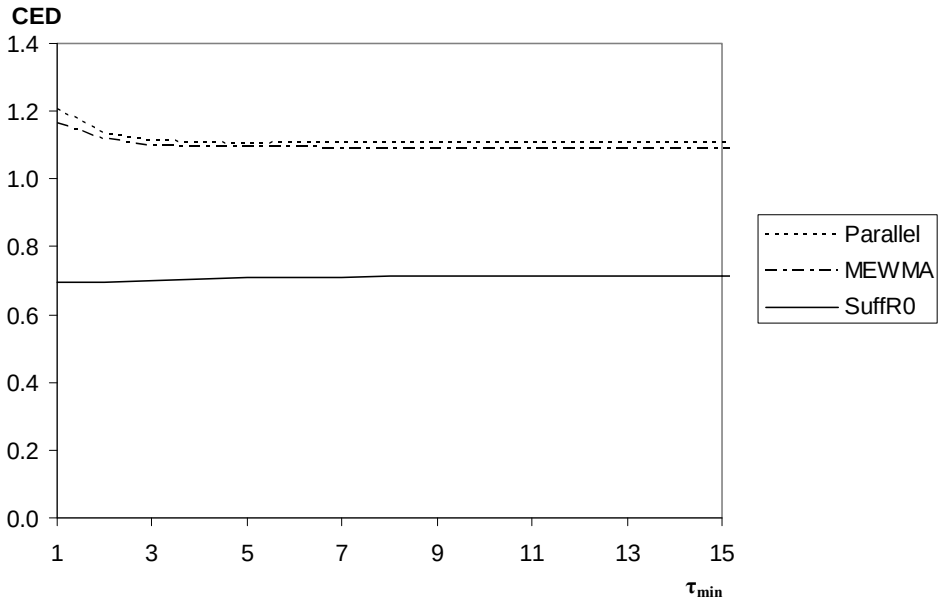

Figure 1. CED(τ_1, τ_2) vs τ_{min} for EWMA based on SuffR^0, EWMA Parallel, and MEWMA, for simultaneous changes, i.e. $\tau_1=\tau_2=\tau_{min}$.

5.4. *Results for changes with a time lag*

We now study the two variables Y_1 and Y_2 in the situation when they change with a known time lag. For the time lag of 1 unit, we find from Corollary 2 that the reduction SuffR^1 should be used. Correspondingly, for a known lag of 5 time units, the SuffR^5 should be used. In Figure 2 we examine the situation when $\tau_2=\tau_1+1$, for $\tau_1 =\tau_{min}$ = 1, 2, ..., 15 and in Figure 3 we examine $\tau_2=\tau_1+5$, for $\tau_1 =\tau_{min}$ = 1, 2, ..., 15. We compare the EWMA method based on the sufficient statistic for the specific situation (lag 1 or lag 5) with MEWMA, a parallel EWMA system, and EWMA based on SuffR^0.

In Figure 2, we can see that EWMA based on the SuffR^1 reduction gives a shorter CED than the other methods for the case when $\tau_2=(\tau_1+1)$.

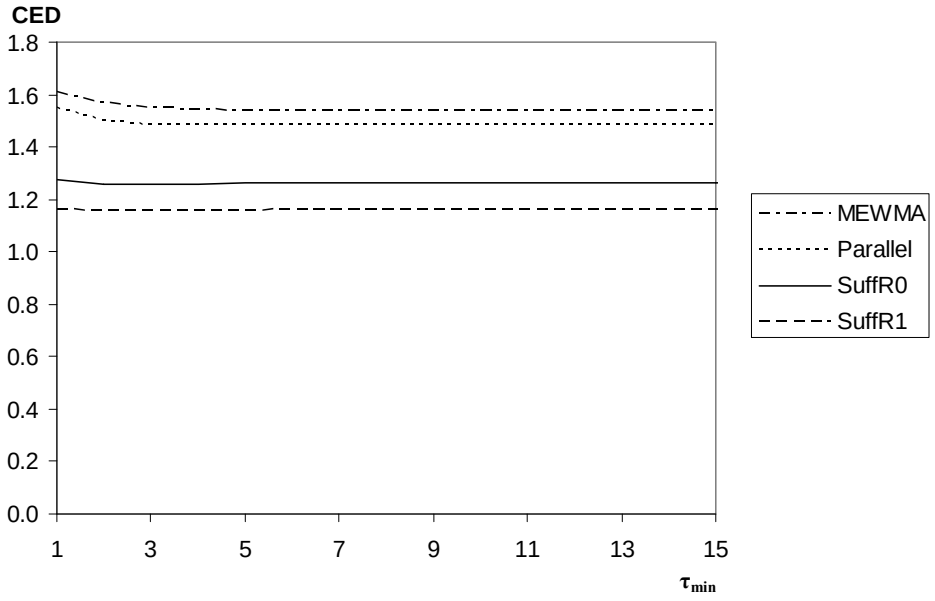

Figure 2. CED(τ_1, τ_2) vs $\tau_1=\tau_{min}$, for $\tau_2=(\tau_1+1)$. Results shown for MEWMA, EWMA Parallel, EWMA based on SuffR^0, and EWMA based on SuffR^1.

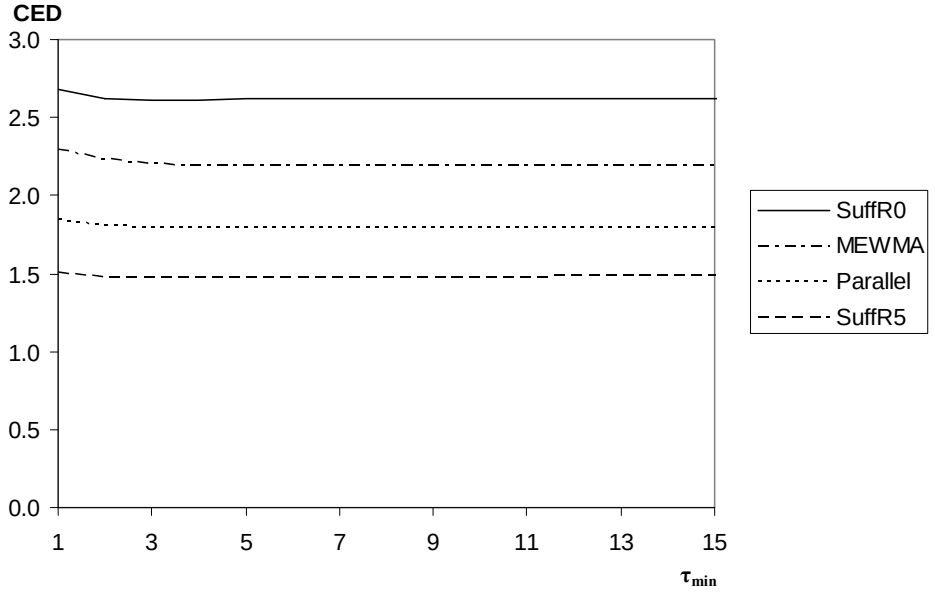

Figure 3. CED(τ_1, τ_2) vs $\tau_1= \tau_{min}$, for $\tau_2=(\tau_1+5)$. Results shown for EWMA based on SuffR^0, MEWMA, EWMA Parallel, and EWMA based on SuffR^5.

In Figure 3 we can see that EWMA based on the $SuffR^5$ reduction has the shortest expected delay.

If we know that only the Y_1 variable can change ($\tau_2=\infty$), then it makes sense to base the surveillance on this variable only, i.e. monitor Y_1 by univariate surveillance. In Figure 4 we see that for $\tau_2=\infty$ the univariate EWMA based on Y_1 is clearly the best alternative. Thus, knowledge considerably improves the CED of the surveillance.

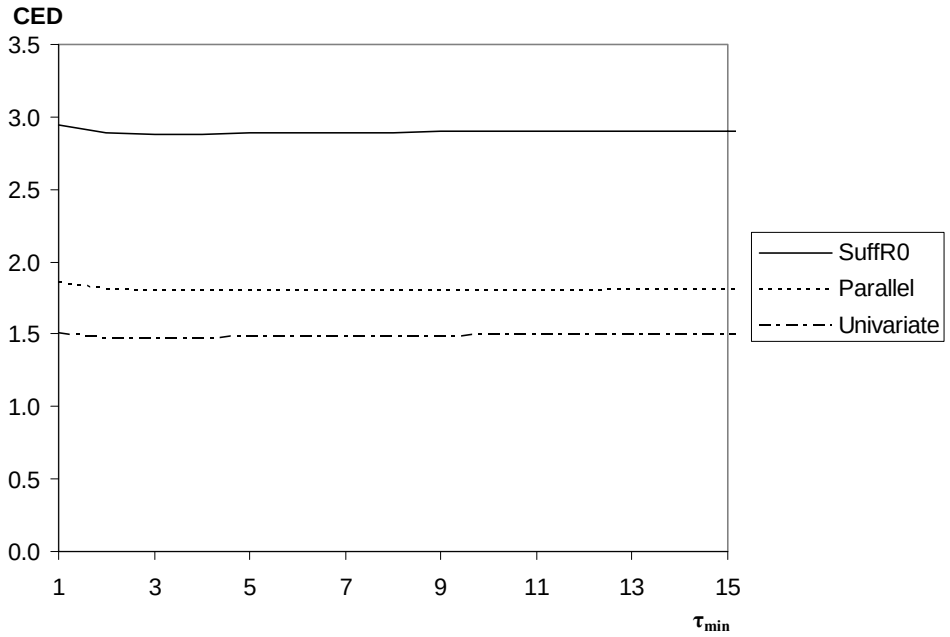

Figure 4. CED(τ_1) vs $\tau_1=\tau_{min}$ for $\tau_2=\infty$. Results shown for EWMA Parallel, EWMA Univariate, and EWMA based on $SuffR^0$.

The conclusion is that for simultaneous changes ($\tau_1=\tau_2$), EWMA based on the $SuffR^0$ reduction gives the shortest delay. This is in accordance with theory, see Wessman, (1998). However, if there is a long time interval between the changes as in Figure 3, or if only one process changes as in Figure 4, the reduction to $SuffR^0$ is not favorable.

6. Discussion

Since many important problems involve several data sources, multivariate surveillance has attracted much interest. It is challenging in many ways. Multivariate surveillance involves statistical theory, practical issues concerning the collection of new types of data, and computational issues such as the implementation of automated methods in large scale surveillance data bases. In this paper the focus has been on the statistical inference aspects and especially the effect of a sufficient reduction of the multivariate surveillance problem. The impact of the relation between the change points is seldom considered. However, here it was demonstrated that the relations between the change points do have a great impact and can be utilized to find efficient methods.

An advantage of the sufficient reduction is that univariate monitoring methods can be used. Properties are often known and optimality is less complicated for univariate methods.

Evaluations are often made by the ARL^1 or the steady state ARL, together with an implicit assumption that all processes change simultaneously. However, if the processes do change simultaneously, there exists a sufficient reduction to a univariate statistic which should be the base for optimal surveillance. Genuinely multivariate problems with different change points should be evaluated by generalized metrics, as suggested in this paper.

According to the sufficiency principle, all conclusions to be drawn should depend only on a sufficient statistic. For simultaneous changes, a univariate optimal accumulation of the information by $SuffR^0$ will result in a jointly optimal surveillance method. We have demonstrated that a considerable improvement can be made by basing the surveillance on the suggested $SuffR^0$ statistic instead of using the Parallel method or the MEWMA.

In the Theorem it is demonstrated, for the exponential family, that a known time lag allows a sufficient reduction. In this situation (i.e. with different change times), the sufficient statistic does not change between two distributions only, and therefore previous optimality results on how to aggregate the information over time cannot be used directly. However, we have demonstrated that for some situations, the method based on a sufficient reduction for the known lag gives the shortest delay to detection compared to a parallel approach or the MEWMA method.

The Theorem shows that there exists a sufficient reduction if the time lag is known. Much statistical inference is derived for a situation with specific assumptions (t-test for $\mu_1=\mu_2$

assumes equal variances, OLS assumes a constant variance over X) and to determine if these assumptions are fulfilled can be challenging. However, if the assumptions are approximately fulfilled, the method will in most cases have good properties. Thus if the time lag is estimated to be approximately 3, the surveillance based on $SuffR^3$ will probably work better than surveillance based on $SuffR^0$. A comparison between Figure 2 and Figure 3 demonstrates that the results are worse for a larger error in the delay than for a smaller. Thus, a close approximation of the true time lag can be assumed to give good results. However, a large error in the assumption on the time lag might result in a less efficient method.

It was also demonstrated – as expected – that in a situation where only one process changes, the performance is considerably improved if this knowledge is utilized in the surveillance procedure.

Acknowledgements

Kjell Pettersson has given constructive comments. The work was supported by the Swedish Emergency Management Agency (grant 0314/206).

References

Andersson, E. (2009). Effect of dependency in systems for multivariate surveillance. *Communications in Statistics. Simulation and Computation* 38:454-472.

Aue, A., Horvath, L. (2004). Delay time in sequential detection of change. *Statistics & Probability Letters* 67:221-231.

Basseville, M., Nikiforov, I. (1993). *Detection of abrupt changes- Theory and application*. Englewood Cliffs: Prentice Hall.

Bersimis, S., Psarakis, S., Panaretos, J. (2007). Multivariate Statistical Process Control Charts: An Overview. *Quality and Reliability Engineering International* 23:517-543.

Bock, D. (2008). Aspects on the control of false alarms in statistical surveillance and the impact on the return of financial decision systems. *Journal of Applied Statistics* 35:213-227.

Cheng, S.W., Thaga, K. (2006). Single Variables Control Charts: an Overview. *Quality and Reliability Engineering International* 22:811-820.

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

Cox, D.R., Hinkley, D.V. (1974). *Theoretical statistics*. London: Chapman and Hall.

Crosier, R.B. (1988). Multivariate Generalizations of Cumulative Sum Quality-Control Schemes. *Technometrics* 30:291-303.

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

Frisén, M. (2009). Principles for Multivariate Surveillance. In: Lenz, H.-J., Wilrich, P.-T., Schmid, W., eds. *Frontiers in Statistical Quality Control 9*. Heidelberg: Physica-Verlag.

Frisén, M., Andersson, E., Schiöler, L. (2009a). Evaluation of Multivariate Surveillance. *Journal of Applied Statistics*.

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

Frisén, M., de Maré, J. (1991). Optimal surveillance. *Biometrika* 78:271-280.

Frisén, M., Sonesson, C. (2006). Optimal surveillance based on exponentially weighted moving averages. *Sequential Analysis* 25:379-403.

Golosnoy, V., Schmid, W., Okhrin, I. (2007). Sequential Monitoring of Optimal PortfolioWeights. In: Frisén, M., ed. *Financial surveillance*. Chichester: Wiley.

Järpe, E. (2000). On univariate and spatial surveillance. Ph.D Thesis. Göteborg University, Göteborg.

Järpe, E. (2001). Surveillance, environmental. In: El-Shaarawi, A., Piegorsh, W.W., eds. *Encyclopedia of Environmetrics*. Chichester: Wiley.

Knoth, S., Schmid, W. (2002). Monitoring the mean and the variance of a stationary process. *Statistica Neerlandica* 56:77-100.

Lee, S., Ha, J., Na, O., Na, S. (2003). The cusum test for parameter change in time series models. *Scandinavian Journal of Statistics* 30:651-739.

Lowry, C.A., Woodall, W.H., Champ, C.W., Rigdon, S.E. (1992). A multivariate exponentially weighted moving average control chart. *Technometrics* 34:46-53.

Lu, C.W., Reynolds Jr, M.R. (1999). EWMA control charts for monitoring the mean of autocorrelated processes. *Journal of Quality Technology* 31:166-188.

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

Okhrin, Y., Schmid, W. (2007). Surveillance of Univariate and Multivariate Nonlinear Time Series. In: Frisén, M., ed. *Financial surveillance*. Chichester: Wiley.

Page, E.S. (1954). Continuous inspection schemes. *Biometrika* 41:100-114.

Reynolds, M.R., Jr, Kim, K. (2007). Multivariate Control Charts for Monitoring the Process Mean and Variability Using Sequential Sampling. *Sequential Analysis* 26:283-315.

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

Roy, S. (1953). On a heuristic method of test construction and ties use in multivariate analysis. *Annals of Mathematical Statistics* 24:220-238.

Ryan, T.P. (2000). *Statistical methods for quality improvement*. 2nd ed. New York: Wiley.

Schiöler, L. (2008). *Explorative analyzis of spatial patterns of influenza incidences in Sweden 1999-2008* (No. 2008:5): Statistical Research Unit, Department of Economics, Göteborg University, Sweden.

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

Shiryaev, A.N. (1963). On optimum methods in quickest detection problems. *Theory of Probability and its Applications.* 8:22-46.

Sonesson, C., Frisén, M. (2005). Multivariate surveillance. In: Lawson, A., Kleinman, K., eds. *Spatial surveillance for public health*. New York: Wiley.

Sun, K., Basu, A.P. (1995). A characterization of a bivariate geometric distribution. *Statistics & Probability Letters* 23:307-311.

Vexler, A., Wu, C. (2009). An optimal retrospective change point detection policy. *Scandinavian Journal of Statistics*.

Wärmefjord, K. (2004). Multivariate quality control and Diagnosis of Sources of Variation in Assembled Products. Licentiat Thesis, Göteborg University, Göteborg.

Wessman, P. (1998). Some Principles for surveillance adopted for multivariate processes with a common change point. *Communications in Statistics - Theory and Methods* 27:1143-1161.

Paper IV

Characterisation of influenza outbreaks in Sweden

Linus Schiöler

Statistical Research Unit, Department of Economics, University of Gothenburg

PO Box 640

SE 405 30 Göteborg, Sweden

E-mail: linus.schioler@statistics.gu.se

Abstract

Aims: The spatial aspect of Swedish seasonal influenza data was investigated and modelled with the main aim of finding patterns that could be useful for outbreak detection, i.e. for detecting an increase in incidence as soon as possible. *Methods:* Quality problems with data on laboratory diagnosed cases (LDI) collected by a number of laboratories and other data were studied. Parametric and nonparametric regression methods were used for estimation of the excepted incidence. Multivariate analysis was used to determine the impact of different spatial components. *Results:* Quality problems were found for all types of data. LDI was found useful for the present aim. No evidence for a geographical pattern was found. It was found that the influenza outbreak started at about the same time in the metropolitan areas and about one week later in the rest of the country. Both parametric and nonparametric regression models are suggested. ***Conclusion:*** **There was a time difference between the outbreaks in the metropolitan areas and the rest of the country. This can be utilised to improve outbreak detection.**

Key Words: Influenza, Sweden, onset of outbreak, statistical models, spatial, monitoring

Background

Influenza is an epidemic disease which causes a significant number of deaths, especially among elderly people and infants, as well as a considerable amount of absenteeism (see for example [1]). Yearly and weekly influenza reports are available from the Swedish Institute for Infectious Disease Control (SMI) at www.smittskyddsinstitutet.se. Statistical models are useful for understanding how the incidence varies. In [2] the problem of modelling influenza data was investigated. A method for predicting the time and height of the peak of the influenza season was proposed in [3].

It is important to detect the onset of the outbreak as soon as possible, in order to be able to allocate the proper resources to the primary care sector and take preventive action. Statistical methods for surveillance increase the chances of early and correct detection. Automatic surveillance systems are now implemented in Sweden [4] and other countries. The three methods implemented in Sweden so far are based on [5-8]. In [7-8] the application of one of the methods to influenza in Sweden is described. This method is applied to the country as a whole. Further development of the methods by incorporating spatial patterns can be beneficial. A known pattern in the spread of the influenza epidemic between regions can be utilised for quicker and more accurate outbreak detection.

Aims

The aim of this paper is to suggest statistical models for incidence at the outbreak of the seasonal influenza. Special emphasis will be put on spatial patterns that could be useful for a surveillance system. There may be a time lag between the outbreaks in different regions of the country, and hence it may be possible to detect an outbreak earlier by considering spatial differences. At a regional level the number of reported influenza cases is small in Sweden, hence some aggregation of data is beneficial. A spatial pattern can be the base for such aggregation.

The modelling of the influenza incidence is important for effective statistical surveillance. Since the variation between years is large, a robust nonparametric or semiparametric model is suitable. A parametric model is needed for simulating data for evaluation purposes. We consider both parametric and semiparametric models.

Material and methods

Swedish data on influenza

In Sweden, several types of data are collected by the SMI during the influenza season. The most established ones are rates of laboratory diagnosed influenza (LDI) and reports by selected physicians on the number of patients with influenza-like illness (ILI). The official reporting from the SMI starts at week 40. Information on the data on influenza in Sweden can be found for example in [9] and at the website of the SMI (www.smittskyddsinstitutet.se). The possibility of collecting data by telephone surveys and self-reporting has also been investigated by the SMI [10-12].

It has been suggested that data on influenza related Internet searches could be used as a proxy for traditional types of data. In [13] a method for using Google's search data is described, and such data on Sweden are available at www.google.org/flutrends/. In [14] search data from a website offering medical advice is used. The website is owned by the Stockholm County Council and is aimed primarily at the residents of Stockholm. Neither of these sources offers spatial information.

The percentage of patients with ILI is a commonly used measure of influenza incidence. However, most regions lack such data for several weeks each year, both on the number of visiting patients and on the number of patients with influenza symptoms. The problem is most evident at the beginning and end of the influenza season. A possible explanation is that medical staff may be less inclined to report cases or perform laboratory testing if there is an expectation that the influenza season has not started or is already over. In [3] it was concluded that the available data on ILI could not be regarded as a good indicator of the incidence. Furthermore, data on the number of patients in each region were not available after the season 04/05. Thus, %ILI for the different regions could not be aggregated in a meaningful way for the later years. The ILI data could therefore not be used for spatial surveillance.

Laboratory diagnosed cases are reported by five viral laboratories and a number of microbiology laboratories. The number of reporting laboratories has increased but varies slightly between the years, as shown in Table I. Although many laboratories have some years missing from their reporting, there are complete data on the period 99/00 to 08/09 for more than half of the regions, including the largest cities (Table II). In [3] it was concluded that LDI is a useful indicator of influenza in Sweden. We will use the sum of the cases with influenza type A and B in our analysis.

The surveillance of spatial clusters of adverse health events has been analysed for example by [15] and [16]. However, in Sweden data are available only for large regions which are not suitable for cluster analysis.

Table I. The number of laboratories having reported confirmed cases to the SMI.

	99-00	00-01	01-02	02-03	03-04	04-05	05-06	06-07	07-08	08-09
Number of laboratories	17	18	20	21	24	24	25	23	25	25

Statistical methods

Spearman's rank correlation was used to investigate the relationship between the time of onset and the coordinates. Linear models were used to further investigate this relationship by incorporating more variables in the analysis.

Data was aggregated to groups and the effect of introducing a time lag between the groups was investigated. We used the root mean square deviation (RMSD), i.e. $\left[\frac{1}{n}\sum_{t=1}^{n}\left(X_1(t)-X_2(t+k)\right)^2\right]^{1/2}$, where $X_1(t)$ and $X_2(t)$ denote the observation t of the first and second group, respectively and k is the time lag. Hence a low value of RMSD is an indicator that the incidences in the two groups agree.

The Swedish influenza incidence will be modelled by a Poisson process with the intensity following an exponential curve, as suggested in [2]. The parametric model is useful to study characteristics of the outbreak pattern and for simulation studies of the properties of surveillance systems.

Since the variation between years is large, a parametric model of the influenza outbreak is of limited use for detecting the outbreak. The major pattern all years is that there is an increase from the onset until a peak is reached, and then a decrease follows. We used nonparametric unimodal regression [17] for the estimation of the expected incidence. This reduces some of the random variation in the available data without any assumption of a parametric model.

Statistical analyses were performed using SAS for Windows version 9.2.

Results

Details of our results are given in the technical report [18]. Here the results are briefly described to support the discussion and the conclusions.

Table II. The total number of laboratory diagnosed influenza cases. Laboratories with data for all years are placed at the top of the table and sorted by median. Laboratories with consistent reporting in later years are placed in the middle and laboratories with inconsistent reporting at the bottom.

	99_00	00_01	01_02	02_03	03_04	04_05	05_06	06_07	07_08	08_09	Median
KS	350	143	215	111	249	282	110	120	247	247	231
Malmö	196	36	149	73	201	359	209	263	158	460	198,5
HS	293	109	178	95	189	252	121	155	185	180	179
Umeå	210	115	195	62	139	165	67	148	98	88	127
Skövde	102	52	140	39	107	184	34	88	15	98	93
Örebro	170	32	83	19	101	76	28	73	55	93	74,5
Göteborg	71	38	47	32	66	41	96	116	146	294	68,5
Falun	65	31	114	20	144	93	44	67	43	101	66
Uppsala	117	47	77	18	34	116	24	36	27	61	41,5
Halmstad	90	18	37	11	42	62	38	52	38	69	40
Karlstad	131	6	40	10	29	73	18	42	13	36	32,5
Kalmar	51	5	36	5	41	91	15	7	25	50	30,5
Linköping	31	5	32	24	23	17	9	16	14	24	20
Uddevalla	66	13	25	9	27	44	12	21	15	18	19,5
Västerås	10	1	9	2	28	29	10	26	4	13	10
Sundsvall		5	51	5	60	46	5	45	51	31	45
Gävle			5	4	15	14	14	20	11	16	14
Karlskrona		9	4	4	15	5	12	2	27	7	
Eskilstuna				2	15	10	2	5	18	15	10
Borås					24	14	7	8	11	21	12,5
Jönköping				12	6	10	24	8	26	11	
Kristianstad						7	27	16	54	21,5	
Lund									26	61	43,5
Helsingborg								15	25	20	
Luleå	22			2	15	14	16		5	6	14
Växjö	32	12	46	7	7	1	1				7
Östersund		9		1	15		1			5	
Kungshamn	5									5	
Trollhättan						2				2	

Spatial pattern

The total number of cases each year is shown in Table II. Laboratories in larger cities tend to report more cases. A large variation between years as well as inconsistent reporting by some laboratories can be noted.

Table III shows the number of weeks to the first laboratory diagnosed influenza case. There is considerable variation between years and also between laboratories. One reason for the latter could be differences in population size. There may also be differences in incidence depending on population characteristics, such as the age distribution, as well as differences in testing policies. The largest cities, Stockholm, Göteborg and Malmö, have generally been

among the first to report cases. Umeå is also generally found among the cities with the earliest reports. Table III also shows the median number of weeks until the cumulative number of LDI cases exceeded 5.

Table III. The number of weeks (from week 40 onwards) to the first laboratory diagnosed influenza case. The regions are sorted with respect to the median week for the first case. The median number of weeks until the cumulative number of LDI cases exceeded 5 is shown in the last column.

	99_00	00_01	01_02	02_03	03_04	04_05	05_06	06_07	07_08	08_09	Median	Median #>5
Göteborg	9	14	14	6	6	6	6	1	0	2	6	14.0
KS	3	14	7	8	5	7	11	10	4	0	7	12.0
HS	3	17	8	13	3	7	8	8	2	6	7.5	12.5
Umeå	3	17	15	12	7	10	5	3	8	7	7.5	14.0
Malmö	3	12	10	15	8	8	13	12	4	5	9	14.0
Borås					6	13	17	14	6	6	9.5	21.0
Skövde	8	14	15	4	5	13	16	8	14	8	10.5	16.5
Lund									11	11	11	15.0
Uppsala	4	14	14	15	8	3	18	11	7	11	11	15.5
Halmstad	9	18	14	17	7	9	14	16	2	2	11.5	18.5
Örebro	10	12	16	18	6	10	20	13	13	8	12.5	18.0
Helsingborg									14	12	13	16.0
Karlstad	6	19	14	15	8	11	17	12	17	3	13	17.0
Luleå	11			12	10	16	17		23	13	13	18.0
Falun	10	17	17	13	8	14	14	12	14	5	13.5	17.0
Jönköping					11	24	14	19	9	13	13.5	20.5
Kristianstad							15	12	18	6	13.5	20.5
Uddevalla	11	16	16	19	7	9	17	16	11	10	13.5	19.0
Sundsvall		21	14	20	8	16	16	11	13	11	14	17.5
Linköping	9	18	18	19	5	10	10	16	16	13	14.5	18.0
Eskilstuna				15	7	18	24	22	15	14	15	18.0
Västerås	12	23	22	20	9	11	18	8	19	3	15	17.0
Kalmar	9	20	16	23	5	16	14	19	15	13	15.5	19.0
Karlskrona			16	16	7	19	13	15	17	11	15.5	22.0
Gävle			18	17	3	16	17	18	10	10	16.5	19.0
Växjö	12	18	16	18	7	25	17				17	21.0
Östersund			19		20	14		18			18.5	23.5

Since the catchment areas of the laboratories differ, the reason that the larger cities reach a larger cumulative sum than the smaller cities could be either that the outbreak occurs earlier in the larger cities or that the probability of a large number is greater for a large population, or a combination of the two. This question will be further studied below.

Spatial analysis often concerns clusters. However, regional data on influenza in Sweden are available only for 25 large regions, which we found unsuitable for standard cluster analysis. Thus, we studied the possible spread to neighbouring areas by analysing how the geographical position indicated by latitude and longitude is associated with the time of the outbreak. Table IV shows the correlations between the coordinates and the number of weeks until the number of LDI cases exceeded 5. None of these correlations differed significantly from zero.

Table IV. Spearman correlation between coordinates and the number of weeks until the number of LDI cases exceeded 5.

	99-00	00-01	01-02	02-03	03-04	04-05	05-06	06-07	07-08	08-09	Median
Latitude	-0.035	0.177	-0.126	-0.261	0.217	-0.129	-0.144	-0.348	0.007	0.134	-0.090
Longitude	-0.021	-0.046	-0.290	-0.431	0.291	-0.200	-0.003	-0.146	-0.133	0.149	-0.177

As the outbreak in general occurred earlier in large cities, we examined classification into two groups: a metropolitan group consisting of Stockholm including Uppsala, Göteborg, Malmö and Umeå, and a locality group consisting of the rest of Sweden. Stockholm, Göteborg and Malmö all have considerably larger populations than the other cities, and they are part of the metropolitan areas as defined in [19]. Uppsala, on the other hand, is more similar in population size to the cities in the locality group. However, the proximity and transport connections to Stockholm make Uppsala suitable to include in the metropolitan group. Moreover, the international airport of Arlanda is situated about halfway between Stockholm and Uppsala. We also included Umeå in the metropolitan group, although the city has a smaller population than the other cities in the group. Umeå is the largest city in the region of Norrland, which comprises about 59 % of the total area and 16% of the population of Sweden. The region's largest hospital is found here. Figure 1 shows the number of LDI cases for each group.

Using Spearman's rank correlation, we found that the pairwise correlations of weekly numbers of LDI cases in Stockholm, Göteborg, Malmö and Umeå were high (correlation coefficient >0.7 for most years). The correlation between Uppsala and the rest of the group was slightly lower but still high enough for it to be reasonable to include Uppsala in the group.

It could be argued that Lund and Borås should also be included in the metropolitan group, due to their proximity to Malmö and Göteborg, respectively. However, the reporting from Borås and Lund was inconsistent. There were also other quality problems associated with the reports from these cities. We chose to exclude them from the metropolitan group.

A multivariate analysis was performed to determine which of the variables year, coordinates and group (metropolitan/locality) had the strongest influence on the time of the onset. To avoid interaction with missing data, only data from laboratories with data for all years were used. Different linear models with the time of the onset as dependent variable were analysed. Year and group were used as qualitative factors and coordinates (latitude and longitude) as continuous variables. We found that the group factor gave the highest partial coefficient of determination apart from year. The latitude and longitude coordinates were not

significant in any of the models. Our conclusion was that there was no strong relation between the coordinates and the time of outbreak.

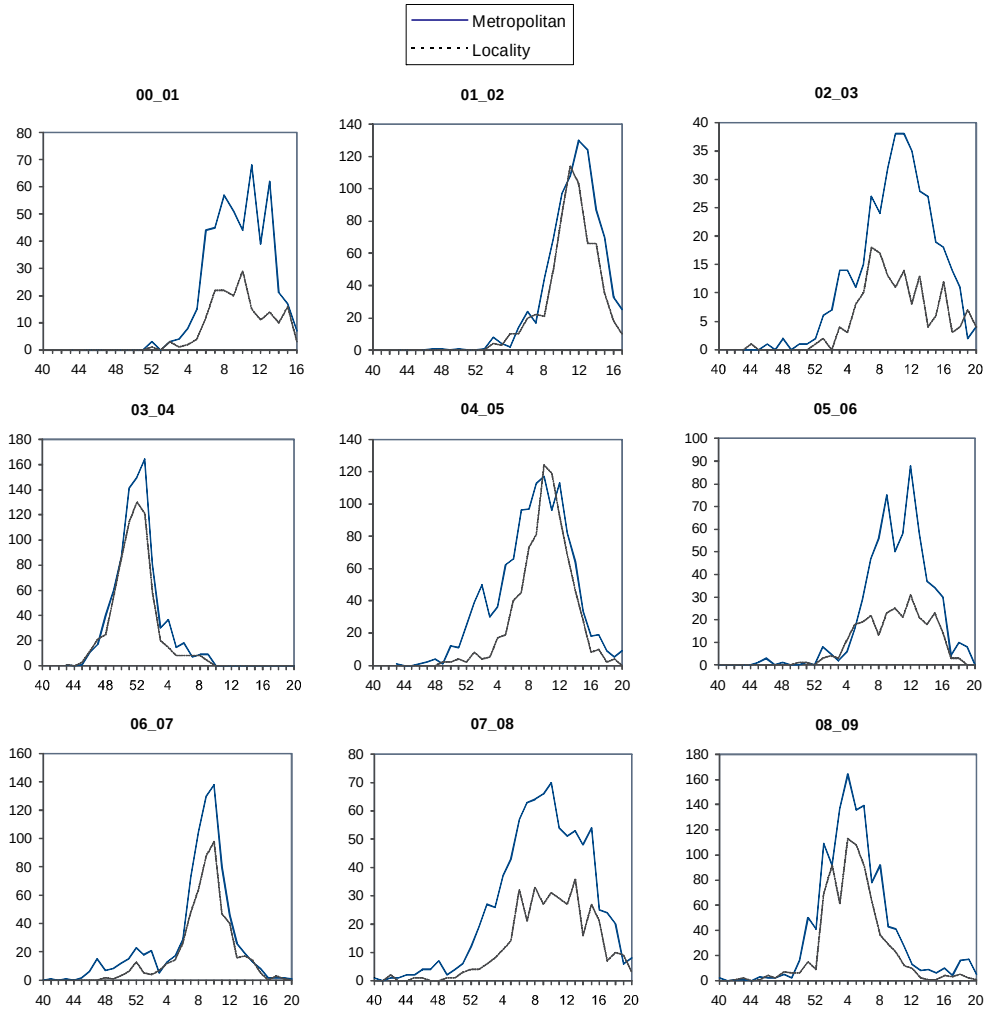

Figure 1. The number of laboratory diagnosed cases for the metropolitan group, Stockholm/Uppsala, Göteborg, Malmö and Umeå (solid line), and the locality group, the rest of Sweden (dotted line).

Time differences in the start of the onset

Table V shows the number of weeks until the cumulative number of LDI cases exceeded 10. This happened first in the metropolitan group in all years except 2002-2003.

Table V. The number of weeks until the cumulative number of LDI cases exceeded 10.

	99_00	00_01	01_02	02_03	03_04	04_05	05_06	06_07	07_08	08_09
Locality	17	16	16	6	13	15	11	12	8	9
Metropolitan	16	14	13	6	10	13	7	6	7	6
Difference	1	2	3	0	3	2	4	6	1	3

Table V suggests that there is a time lag between the two groups. Additional analyses on each influenza season were performed to see which shift in time would make the incidences in the metropolitan and locality areas more alike. We calculated the total root mean square deviation, including all influenza seasons in the calculation. Different time lags between the two groups were examined. In the presence of a time lag we would expect the lowest RMSD for the correct value of the lag. Since our primary interest is the outbreak, we used only the observations from the start and until the number of observed cases in the metropolitan group had exceeded 15. The results are shown in Table VI. The total RMSD was lowest for a lag of one week.

Table VI. Root mean square deviation between the metropolitan and locality groups.

Lag	RMSD
0	5.75
1	5.15
2	6.95

The uptake area of each laboratory is not known and therefore population size cannot be used in the analysis. A larger population means that a fixed number of cases will be exceeded earlier, even if the incidences are the same. The number of cases was larger for the metropolitan group. The median number of cases at the peak of the incidence was 123.5 for the metropolitan group and 105.5 for the locality group, a ratio of 1.17. To study the effect of the difference in size, we adjusted the size of the groups in the parametric model defined below and compared the time it took for the cumulative sum to exceed 5. The resulting time difference after the adjustment was about one day. Thus, a difference in population size of this magnitude could not be seen as the full explanation for the observed difference in the time of outbreak.

Parametric models of the expected incidence

A parametric model is useful to describe details of the outbreak. In order to make a simulation study of the properties of a surveillance method, some sort of parametric model is also needed. In [7] the model

$$\mu(t) = \begin{cases} \mu_0, & t < \tau \\ \exp(\beta_0 + \beta_1(t - \tau + 1)), & t \geq \tau \end{cases},$$

where τ denotes the time of the onset, is used for a typical curve of the total number of LDI cases in the whole of Sweden. The constant phase, μ_0, was roughly estimated to $\mu_0 = 1$ from Swedish LDI data for eight years. The model was estimated from the incidence in the season 03/04, when the outbreak was neither particularly severe nor particularly mild. The estimates of the parameters were $\beta_0 = -0.26$ and $\beta_1 = 0.826$.

By the results above we have that the locality and metropolitan groups each had about half the number of cases in Sweden as a whole and an approximate time lag between them of about one week. Thus, the relation between the incidences of the total (T), metropolitan (M) and locality (L) areas can be expressed by

$$\mu_T(t) = \mu_M(t) + \mu_L(t) = \begin{cases} \mu_0, & t < \tau_M \\ \exp(\beta_0^* + \beta_1^*(t - \tau_M + 1)) + \mu_0 / 2, & \tau_M \leq t < \tau_L \\ \exp(\beta_0^* + \beta_1^*(t - \tau_M + 1)) + \exp(\beta_0^* + \beta_1^*(t - \tau_L + 1)) & t \geq \tau_L \end{cases}$$

where $\tau_L = \tau_M + 1$ and $\mu_0 = 1$. The parameters $\beta_0^* = -0.62$ and $\beta_1^* = 0.826$ give a good approximation of the model for the total incidence above. This curve fitted well to the data for the same season (03/04) for some values of the starting time. It also fitted rather well for some other seasons, while a good fit for all seasons could not be expected due to the marked differences between the seasons.

Nonparametric and semiparametric models of the expected incidence

Due to the limited quality and the variation between years, the parametric model is unsuitable for inference. The interaction between the estimates of the start and slope of the outbreak is another weakness of parametric models. The use of order restrictions for modelling outbreaks is suggested in [20], where it is assumed that the incidence is constant up to some starting point and then non-decreasing. A similar assumption is used in [3], where the time of onset and the slope are used for predicting the time and height of the peak in influenza incidence. The time difference between the (interpolated) time points when the total number of LDI cases

in Sweden exceeds 30 and 10, respectively, is used as an indicator of the slope. We applied these techniques to the aggregated data but used the time difference between 15 and 5, since each of the groups accounts for about half of the total number of cases in Sweden. We found no significant difference between the slopes of the metropolitan and locality groups.

The nonparametric model by order restriction can be combined with the Poisson distribution to a semiparametric model. In [8] a semiparametric method of surveillance is applied to Swedish LDI data for the country as a whole. Figure 2 shows the alarm statistic of the method applied to the metropolitan and locality groups. The metropolitan group had a tendency to an earlier increase than the locality group. Thus, an earlier alarm or first warning can be expected here.

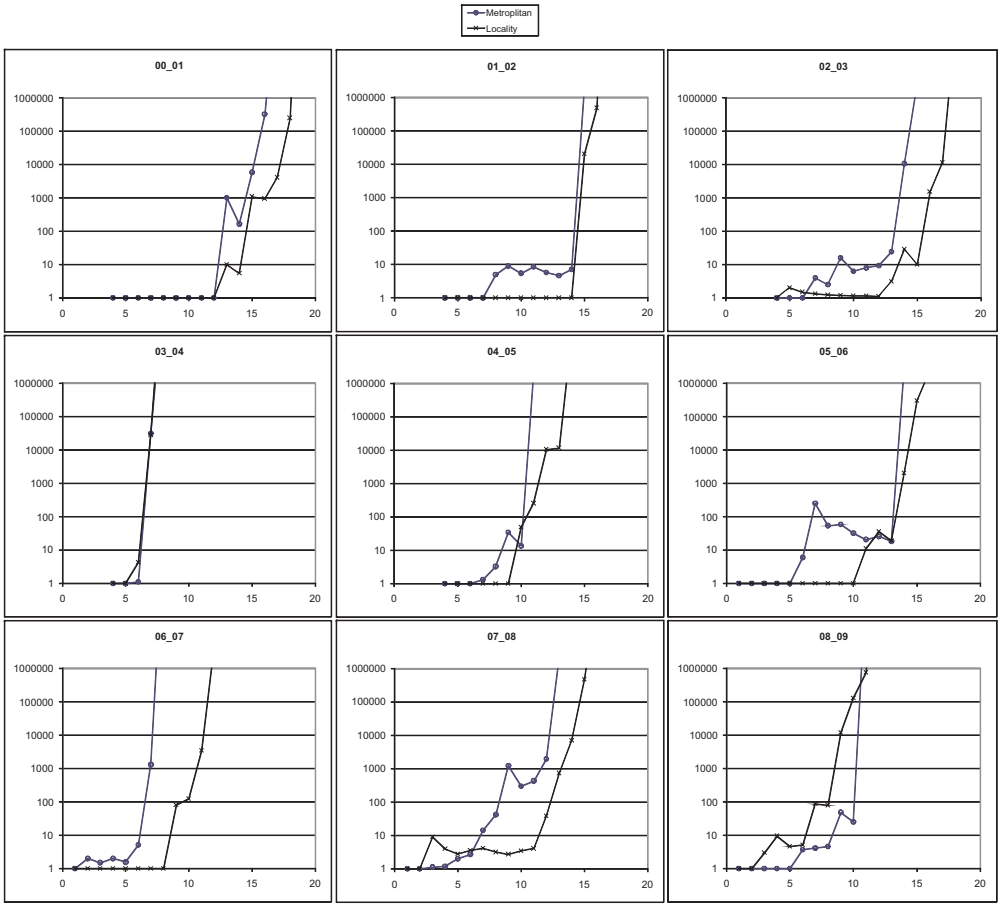

Figure 2. OutP alarm statistics for the metropolitan group (dots) and the locality group (crosses).

Discussion

The surveillance of infectious diseases such as influenza has drawn much attention recently. We analysed the spatial aspect of Swedish influenza data with the main aim of finding patterns that could be useful for statistical surveillance of the outbreak, i.e. for detecting an increase in incidence as soon as possible.

In Sweden, several types of data are collected during the influenza season. The most established ones are data on laboratory diagnosed cases (LDI), collected by a number of laboratories, and cases of influenza-like illness (ILI), collected by a number of selected physicians. Quality problems were found for both types of data but were most severe for ILI. A potential problem with LDI data is that policies regarding testing may differ between administrative areas. Hospitals conducting research on influenza may also be more inclined to perform testing. The differences in population size between the catchment areas of the laboratories may also constitute a problem. The number of cases can be expected to be greater for laboratories serving large populations. Thus, one has to be careful with drawing conclusions regarding the incidence from the number of confirmed cases, since a higher number of cases can be the result of both a higher incidence and a larger population. The varying number of reporting laboratories may also be a problem, particularly when using a surveillance method that relies on a baseline to distinguish between the epidemic and non-epidemic phases. However, the fact that primarily smaller laboratories are inconsistent in their reporting lessens this effect.

In [7-8] it has been shown that Swedish influenza data can be useful for surveillance. By combining results from different parts of the country in an efficient way, inference regarding the outbreak in the country as a whole might be performed more efficiently. We found that there was a time lag between the metropolitan and locality areas. This can be potentially useful for faster and more reliable detection of the outbreak.

Spatial patterns such as those based on geographical coordinates were examined. We found no evidence for a relation between the time of the onset of the outbreak and a location to the north/south or east/west. We found that in the major cities, Stockholm (including Uppsala), Göteborg, Malmö and Umeå, the onset of the influenza outbreak seemed to occur earlier than in the rest of the country. Analysis with respect to the variables coordinates, group (metropolitan/locality) and year revealed that year and group was the most important as concerns the time of the onset of the outbreak. These metropolitan regions all have major airports nearby, and commuting is common.

The properties of the metropolitan and the locality groups were analysed by studying the time at which a certain incidence was reached, the similarity between lagged variables, and graphs of the incidence and alarm statistic at the onset. Although the variation between years was quite large, a difference of one week between the metropolitan and locality groups was a good approximation for most years. There are a number of factors that could contribute to the difference in influenza incidence between regions. Temperature and humidity affect the transmission of influenza virus [21]. This may be a factor in Sweden due to its diverse climate. However, we found no influence of the geographical coordinates, which are of course correlated with climate variables. Air travel has been found to have significant effect on the spread of influenza in the USA [22]. It is thus probable that major cities with well-developed means of transport may have an earlier outbreak than smaller cities.

Stochastic models for influenza incidence are needed for many purposes. An earlier study [2] has found that the Poisson distribution fits well to data at the onset of the outbreak. In this paper, parametric exponential regression models were suggested for the metropolitan and locality groups separately. As for the incidence slope at the onset, no evidence was found for a difference between the two groups. These parametric models are useful to generate data for simulation and for enhancing understanding. The variation in incidence between the years is large. Therefore, a nonparametric or semiparametric approach would be more suitable. For surveillance purposes, we suggest using a robust nonparametric regression model with order restriction.

Conclusion

Geographical coordinates such as the location to the north/south or east/west had little influence on the time of the onset of the influenza outbreak. The dominating spatial pattern was that for the major cities, Stockholm (including Uppsala), Göteborg, Malmö and Umeå, the onsets of the outbreak occurred earlier than in the rest of the country. A time difference of about one week between the metropolitan and locality groups was observed.

An exponential regression, with the same slope for the metropolitan and locality groups, fitted well to the data at the onset of the outbreak. However, the parameters differed much between years, and for surveillance purposes we therefore recommend a nonparametric regression with a constant phase before the onset and a monotonically increasing phase from the onset onwards.

Acknowledgements

The author is grateful to Marianne Frisén, Eva Andersson and Kjell Pettersson for constructive comments on the statistical analysis. The data were made available to us by the Swedish Institute for Infectious Disease Control, and we are grateful for discussions about the aims and the data quality. The research was supported by the Swedish Civil Contingencies Agency (grant 0314/206).

References

1. Molinari N-AM, Ortega-Sanchez IR, Messonnier ML, Thompson WW, Wortley PM, Weintraub E, et al. The annual impact of seasonal influenza in the US: Measuring disease burden and costs. Vaccine. 2007;25(27):5086-96.
2. Andersson E, Bock D, Frisén M. Modeling influenza incidence for the purpose of on-line monitoring. Stat Methods Med Res. 2008;17:421-38.
3. Andersson E, Kuhlmann-Berenzon S, Linde A, Schiöler L, Rubinova S, Frisén M. Predictions by early indicators of the time and height of yearly influenza outbreaks in Sweden. Scand J Public Health. 2008;36:475-82.
4. Cakici B, Hebing K, Grünewald M, Saretok P, Hulth A. CASE –a framework for computer supported outbreak detection. . BMC Med Inform Decis Mak. 2010;10(14).
5. Farrington CP, Andrews NJ, Beal AD, Catchpole MA. A statistical algorithm for the early detection of outbreaks of infectious disease. J R Statist Soc A. 1996;159:547-63.
6. Kulldorff M. A spatial scan statistic. Comm Stat Theor Meth. 1997;26(6):1481-96.
7. Frisén M, Andersson E. Semiparametric surveillance of monotonic changes. Sequential Analysis. 2009;28(4):434-54.
8. Frisén M, Andersson E, Schiöler L. Robust outbreak surveillance of epidemics in Sweden. Stat Med. 2009;28(3):476-93.
9. Brytting M, Stivers M, Dahl H, Serifler F, Linde A, Rubinova S. Annual Report july 2007-june 2008 : The National Influenza Reference Center. Solna: Swedish Institute for Infectious Disease Control; 2006.
10. Payne L, Kühlmann-Berenzon S, Ekdahl K, Giesecke J, Högberg L, Penttinen P. 'Did you have flu last week?' A telephone survey to estimate a point prevalence of influenza in the Swedish population. Eurosurveillance. 2005;10(12):241-4.
11. Brytting M, Stivers M, Dahl H, Serifler F, Linde A, Rubinova S. Annual Report July 2006 - June 2007: The National Influenza Reference Center. Solna: Swedish Institute for Infectious Disease Control; 2006.
12. Bexelius C, Merk H, Sandin S, Ekman A, Nyrén O, Kühlmann-Berenzon S, et al. SMS versus telephone interviews for epidemiological data collection: feasibility study estimating influenza vaccination coverage in the Swedish population. Eur J Epidemiol. [10.1007/s10654-008-9306-7]. 2009;24(2):73-81.

13. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature. [10.1038/nature07634]. 2009;457(7232):1012-4.
14. Hulth A, Rydevik G, Linde A. Web Queries as a Source for Syndromic Surveillance. PLoS ONE. 2009;4(2):e4378.
15. Kulldorff M. Prospective time periodic geographical disease surveillance using a scan statistic. J R Statist Soc A. 2001;164(1):61-72.
16. Sonesson C. A CUSUM framework for detection of space-time disease clusters using scan statistics. Stat Med. 2007;26:4770-89.
17. Frisén M. Unimodal regression. The Statistician. 1986;35(4):479-85.
18. Schiöler L. Modelling the spatial patterns of influenza incidence in Sweden. Gothenburg: Statistical Research Unit, Department of Economics, University of Gothenburg, Sweden; 2010. Report No.: 2010:1.
19. Statistiska centralbyrån. Geografin i statistiken - regionala indelningar i Sverige Statistiska centralbyrån; 2005. Report No.: 2005:2.
20. Frisén M, Andersson E, Pettersson K. Semiparametric estimation of outbreak regression. Statistics. 2010;44(2):107-17.
21. Lowen AC, Mubareka S, Steel J, Palese P. Influenza Virus Transmission Is Dependent on Relative Humidity and Temperature. PLoS Pathogens. 2007;3(10):e151.
22. Brownstein JS, Wolfe CJ, Mandl KD. Empirical Evidence for the Effect of Airline Travel on Inter-Regional Influenza Spread in the United States. PLoS Medicine. 2006;3(10):e401.

Paper V

Multivariate outbreak detection

BY LINUS SCHIÖLER[1] and MARIANNE FRISÉN[1]

University of Gothenburg

On-line monitoring is needed to detect outbreaks of diseases like influenza. Surveillance is also needed for other kinds of outbreaks, in the sense of an increasing expected value after a constant period. Information on spatial location or other variables might be available and may be utilized. We adapted a robust method for outbreak detection to a multivariate case. The relation between the times of the onsets of the outbreaks at different locations (or some other variable) was used to determine the sufficient statistic for surveillance. The derived maximum likelihood estimator of the outbreak regression was semi-parametric in the sense that the baseline and the slope were non-parametric while the distribution belonged to the exponential family. The estimator was used in a generalized likelihood ratio surveillance method. The method was evaluated with respect to robustness and efficiency in a simulation study and applied to spatial data for detection of influenza outbreaks in Sweden.

1. Introduction

On-line surveillance is used to give an alert signal as soon as possible after an important change has occurred. Overviews of the inferential issues in surveillance are given by Lai (1995), Woodall and Montgomery (1999), Ryan (2000), Frisén (2003), Frisén (2009) and others.

Here we will consider the detection of an outbreak, defined as a change from a (possibly unknown) baseline to a monotonically increasing (or decreasing) regression. Other definitions of outbreaks are discussed in Section 0.

The motive for this study was the spatial surveillance of influenza outbreaks. The detection of outbreaks of epidemiological diseases is an important area of on-line surveillance. Surveillance in public health is reviewed by for example Sonesson and Bock (2003), Lawson and Kleinman (2005), Woodall (2006), Shmueli and Burkom (2010), and Kass-Hout and Zhang (2010). By monitoring incidences, outbreaks of reoccurring diseases may be detected, for example the yearly influenza epidemic. Such monitoring is also useful to detect new diseases, such as SARS, avian flu and swine influenza, as well as effects of bioterrorism. Early detection of the onset of an outbreak is useful in order for health authorities to act timely and also for the planning of health care. Epidemics, such as influenza, are for several reasons very costly to society and it is therefore of great value to monitor the epidemic period in order to properly allocate medical resources (Andersson et al. (2008b)). A semi-parametric method for detecting the onset of a monotonic increase was suggested for univariate surveillance by Frisén and Andersson (2009). It was successfully applied to the incidence of influenza in Sweden as a whole by Frisén et al. (2009).

As information on the incidence in different regions of the country is available, we will here generalize the univariate method to utilize this information. Spatial surveillance is a special case of multivariate surveillance, as pointed out for example by Sonesson and Frisén (2005) and Joner Jr. et al. (2008). The relation between different variables (here locations) is important in the monitoring of the onset of the outbreak. We will use information from a study by Schiöler (2010) on the spread of influenza in Sweden. The spreading pattern is described in Section 6.1. We will investigate how information on time lags in the onset at different locations should be used in an outbreak surveillance system. Another case there a time lag

[1] Supported by the Swedish Civil Contingencies Agency (grant 0314/206).
Key words and phrases. Exponential family, Generalised likelihood, Ordered regression, Spatial data, Surveillance.

might be relevant is when you have an early but rough indicator which might be combined with a later and more accurate one,. In Hulth et al. (2009) and Ginsberg et al. (2009) it was shown that data of search patterns on the Internet could be used as a proxy for influenza incidence. Ginsberg, et al. (2009) found that the lag in reporting was about one day compared to between one and two weeks for traditional CDC-data. The method suggested in this article may possibly be useful also for situations like that one, where the lag is in the reporting rather than in the onset of the outbreak at the various locations.

In Section 2, we will specify univariate and multivariate models for outbreaks. In Section 3, we will derive a sufficient reduction of the data for multivariate outbreak situations. Sufficient reduction for detection of step changes was earlier derived by Frisén et al. (2010c) but here it is derived for detection of gradual outbreaks. In Section 4, we will discuss general approaches of how multivariate surveillance can be constructed from univariate surveillance, and construct a simple multivariate outbreak detection method, based on the univariate method by Frisén and Andersson (2009). In this section, we will also derive the recommended method. This is done by deriving the maximum likelihood estimators based on the multivariate monotonicity restrictions and using these in a generalized likelihood ratio method. In Section 5, we evaluate the suggested method by a simulation study, where properties like predictive value and robustness are examined. The robustness is important since you never can expect assumptions to be exactly fulfilled. In the comparison with other methods we will use the evaluation metrics suggested by Frisén et al. (2010b) for multivariate surveillance. In Section 6, the method is applied to data for several influenza seasons in Sweden, and the efficiency of the suggested multivariate outbreak detection method is demonstrated. Concluding remarks are given in the final section.

2. Specification of the outbreak model

At each time point, t, a new observation is made on a process **Y**. We want to detect the change from one state to another as soon as possible after it has occurred, in order to give warnings and to take corrective actions.

2.1. Univariate outbreak

In Andersson et al. (2008a) Swedish influenza data from six seasons (2001–2007) were analyzed, and it was suggested that a non-parametric approach based on monotonicity restrictions (the outbreak regression) should be used. It was also suggested that the outbreak could be modeled using a Poisson distribution for the incidence. The parameter $\lambda(t)$ of the distribution at time t has a constant value λ_0 before the outbreak but depends on time after the onset of the outbreak. We will use τ to denote the unknown time of the onset. Thus

$$\lambda(t) = \begin{cases} \lambda_0, & t < \tau \\ \lambda_{t-\tau+1}, & t \geq \tau \end{cases}$$

with $\lambda_0 \leq \lambda_1 \leq \lambda_2 \leq ... \leq \lambda_s$. The aim at decision time s is to determine whether or not the outbreak has started yet, thus if $\tau \leq s$ or $\tau > s$. The state at the outbreak is characterized by a monotonically increasing expected incidence.

The situation where the regression is constant at first and then monotonically increasing will be called “outbreak regression”.

2.2. Multivariate outbreak

In multivariate surveillance the process under surveillance is a p-variate vector, denoted by $\mathbf{Y} = \{\mathbf{Y}(t), \quad t = 1, 2, ...\}$, where $\mathbf{Y}(t) = \{Y_1(t), Y_2(t), ..., Y_p(t)\}$. The components of the vector represent, for example, the incidence of a disease at p different locations. Each component has the same properties as $\lambda(t)$ described in Section 2.1. The time of the onset may differ for the components and will be denoted τ_i for component i. At decision time s, we base the decision whether an outbreak has occurred or not on the available information, $\mathbf{Y}^s = \{\mathbf{Y}(1), \mathbf{Y}(2) ... \mathbf{Y}(s)\}$.

When several processes are observed, knowledge about the relation between the times of the onsets of the outbreaks is essential. Different methods are suitable for different relations. The aim is to detect an outbreak in any of the processes, which means that we aim at detecting the first one. The time τ_i of the onset of the outbreak of process Y_i may not be the same for all i=1,...p. The relation between the times is important. We will concentrate on the case of a known time lag. This can be the case for spatial data and data from several sources (possibly including proxy data). The case where the lag is misspecified is examined in Section 5.5. For notational convenience we order the processes according to which changes first, so that $\tau_1 \leq ... \leq \tau_p$, and denote the time lag for process Y_i by q_i, where $q_1=0$ and $q_i = \tau_i - \tau_1$ for i=2,...,p. The case where the onsets are simultaneous, that is $\tau_i = \tau$ for i=1,...p, is of special interest. In this case $q_i=0$ i=1,...p. We denote this by lag=0. In numerical examples and applications we will also use the special cases of two processes with $q_2=1$ or $q_2=2$. We denote this by lag=1 and lag=2, respectively.

We assume that the distributions of the processes all belong to the one-parameter exponential family. In the application to influenza data in Section 6, the Poisson distribution is relevant.

If a parametric shape of the outbreak pattern is known, this should be used to increase efficiency. However, we do not assume a parametric outbreak pattern here. Instead, we assume that the different processes are identically distributed except for the time of the onset.

3. Sufficient reduction at multivariate outbreaks

In Frisén, et al. (2010b) it was demonstrated that the relation between the change points of the different processes is very important, since it affects the properties of different surveillance methods in different ways. In simple examples, it was demonstrated that a method which is optimal for simultaneous changes is inefficient in other cases. Thus, any knowledge on the change points should be utilized. A sufficient reduction will not reduce the information and still allows a joint solution to the full surveillance problem. It is of special interest to study a simultaneous outbreak at all locations and also a time lag in the onset of the outbreaks. Robustness when the time lag is only approximately known is studied in Section 5.5.

3.1. Simultaneous change at all locations

Many evaluations of multivariate surveillance methods are made by the zero-state ARL (see Section 5.3) where the change occurs at the start. When all processes change at the start it follows that they change simultaneously.

Wessman (1998) and Frisén, et al. (2010c) demonstrated that if all processes have the same change points, i.e. $\tau_1 = \tau_2 = ... \tau_p = \tau$, then the univariate vector of partial likelihood ratios, $\{L(s,t), t=1,...s\}$ where $L(s,t) = f(Y; \tau = t \leq s) / f(Y; \tau > s)$ is sufficient for the sequence of distributional families. Thus, in order to monitor a simultaneous fully specified change in distribution, it is possible to construct a univariate surveillance procedure based on the sufficient sequence of likelihood ratios. Zhou et al. (2010) used this result for the simultaneous

shifts of mean and variance in a normal distribution. For the case with no lag between the change points of two processes (lag=0), the sufficient statistic is denoted by SuffR0. We will use this notation in the application of spatial surveillance of Swedish influenza outbreaks. In this case, SuffR0 corresponds to the total incidence in the country as a whole. The statistic OutbreakPSuffR0 of the method in the application is hence equivalent to the statistic of the univariate surveillance of influenza in Sweden reported in Frisén and Andersson (2009) and Frisén, et al. (2009).

3.2. Changes with a time lag between locations

Järpe (2000) studied the case of a known time lag for independent normal distributions with equally sized shifts in the expected value at the change points and demonstrated that a sufficient reduction to univariate surveillance exists. Frisén, et al. (2010c) studied the case of changes in the general one-parameter exponential family (including the Poisson distribution) but also only for step changes. Different levels of the parameter before the change as well as differences in shift size were considered.

The earlier results on sufficiency for the detection of a step change cannot be used directly for outbreak detection, since we are interested in detecting a change from a constant level to a monotonically increasing one rather than a sudden shift. Here, we study the case where each process Y_i increases monotonically from the onset of the outbreak τ_i and onwards and there is a known time lag between the onsets of each process. The indices of the observation vectors $\{\mathbf{y}_1, \mathbf{y}_2, \ldots \mathbf{y}_p\}$ are ordered according to ascending time lag, i.e. the change occurs first in $\mathbf{Y}_1$. Theorem 1 shows that a sufficient reduction to a univariate statistic exists for the situation with different outbreak times, and in Example 1 (after Theorem 1 and its proof) the theorem is illustrated for a simple case. A numerical illustration is given in Example 2 in Section 4.6.

Theorem 1: For p processes $Y_1, Y_2, \ldots, Y_p$ which all belong to the one-parameter exponential family and which are independent and identically distributed, conditional on the change points and time lags (independent over time as well as across processes), there exists a sufficient reduction of the set of observation vectors to a univariate statistic for the detection of outbreaks with equal (but possibly unknown) parameter values from the onset of the outbreak when the changes occur with known time lags ($q_1=0, q_2, q_3, \ldots q_p$) where $q_i=\tau_i - \tau_1$. A sufficient statistic for inference on the first onset τ_1 is the sequence

$$\sum_{i\in I_t} Y_i(t+q_i)\ \text{t=1,...s, where}\ I_t=\{i: q_i \le s-t, 1\le i\le p\}.$$

This is true both for the situation when the time of change is fixed but unknown and for a stochastic time of change.

Proof: Since the observations are independent given the values of the change points, the distribution can be written as a product. We will first consider a fixed but unknown value of τ_1. The likelihood expressions for the one-parameter exponential family can be written as

$$f(Y;\tau_1 \le s) =$$
$$\exp\left\{\sum_{i=1}^{p}\sum_{t=1}^{\min(\tau_i-1,s)}\left[y_i(t)(\varphi_0)+g(\varphi_0)+h(y_i(t))\right]+\sum_{i=1}^{p}\sum_{t=\tau_i}^{s}\left[y_i(t)(\varphi_{t-\tau_i+1})+g(\varphi_{t-\tau_i+1})+h(y_i(t))\right]\right\}$$

and

$$f(Y;\tau_1 > s) = \exp\left\{\sum_{t=1}^{s}\sum_{j=1}^{p}\left[y_j(t)(\varphi_0) + g(\varphi_0) + h(y_j(t))\right]\right\}.$$

Thus, we have the log likelihood ratio

$$\log\frac{f(Y;\tau_1 \le s)}{f(Y;\tau_1 > s)} = \sum_{i=1}^{p}\sum_{t=1}^{\min(\tau_i-1,s)}\left[y_i(t)(\varphi_0) + g(\varphi_0) + h(y_i(t))\right]$$

$$+\sum_{i=1}^{p}\sum_{t=\tau_i}^{s}\left[y_i(t)(\varphi_{t-\tau_i+1}) + g(\varphi_{t-\tau_i+1}) + h(y_i(t))\right] - \sum_{t=1}^{s}\sum_{i=1}^{p}\left[y_i(t)(\varphi_0) + g(\varphi_0) + h(y_i(t))\right]$$

$$= \sum_{i=1}^{p}\sum_{t=\tau_i}^{s}\left[y_i(t)(\varphi_{t-\tau_i+1}) - y_i(t)(\varphi_0) + g(\varphi_{t-\tau_i+1}) - g(\varphi_0)\right]$$

$$= \sum_{i=1}^{p}\sum_{t=\tau_1+q_i}^{s}\left[y_i(t)(\varphi_{t-(\tau_1+q_i)+1} - \varphi_0)\right] + z(\varphi_0,...,\varphi_{s-\tau_1+1})$$

$$= \sum_{i=1}^{p}\sum_{t=\tau_1}^{s-q_i}\left[y_i(t+q_i)(\varphi_{t-\tau_1+1} - \varphi_0)\right] + z(\varphi_0,...,\varphi_{s-\tau_1+1})$$

$$= \sum_{t=\tau_1}^{s}\sum_{i\in I_t}\left[y_i(t+q_i)(\varphi_{t-\tau_1+1} - \varphi_0)\right] + z(\varphi_0,...,\varphi_{s-\tau_1+1})$$

$$= \sum_{t=\tau_1}^{s}(\varphi_{t-\tau_1+1} - \varphi_0)\sum_{i\in I_t}\left[y_i(t+q_i)\right] + z(\varphi_0,...,\varphi_{s-\tau_1+1}),$$

which depends on the observations only through the statistic in the theorem. The likelihood ratio is sufficient for the problem, and hence the statistic is sufficient. This completes the proof when τ_1 is fixed but unknown.

If τ_1 is stochastic with some distribution g(t), then the density of Y can be written:

$$f(Y) = \sum_{t=1}^{\infty} g(t) f(Y \mid \tau_1 = t),$$

which is a function of $f(Y \mid \tau_1 = t),$ and hence the arguments above can be used to show that the statistic in Theorem 1 is sufficient for the problem also in this case.■

Since any one-to-one function of a sufficient statistic is sufficient, the sequence

$$\sum_{i\in I_t} Y_i(t+q_i)/\mid I_t \mid : t = 1,...,s,$$

where $\mid I_t \mid$ denotes the cardinality of I_t, is also sufficient. This transformed statistic is useful when dealing with the monotonicity restrictions of the outbreak regression, since this statistic preserves the monotonicity properties.

When we have two processes we will use a simpler notation, SuffRq(s,t)= $\sum_{i\in I_t} Y_i(t+q_i)/\mid I_t \mid : t = 1,...,s$,where q is the lag between the two processes.

EXAMPLE 1

For two processes Y_1 and Y_2 with time lag q=1, the index set is $I_t = \{i : q_i \le s-t, 1 \le i \le p\}$.

For s=1 we have $I_1 = \{i : q_i \le 0, 1 \le i \le 2\} = \{1\}$.

For s=2 we have $I_1 = \{i : q_i \le 1, 1 \le i \le 2\} = \{1,2\}$ and $I_2 = \{i : q_i \le 0, 1 \le i \le 2\} = \{1\}$.

For s=3 we have $I_1 = \{i : q_i \le 2, 1 \le i \le 2\} = \{1,2\}$, $I_2 = \{i : q_i \le 1, 1 \le i \le 2\} = \{1,2\}$ and $I_3 = \{i : q_i \le 0, 1 \le i \le 2\} = \{1\}$. Hence, the sufficient reduction is $\left\{\sum_{i=1} Y_i(t) : t = 1\right\} = \{Y_1(1)$ at s=1, $\left\{\sum_{i\in I_t} Y_i(t+q_i) : t = 1,2\right\} = \left\{\sum_{i\in\{1,2\}} Y_i(1+q_i), \sum_{i\in\{1\}} Y_i(1+q_i)\right\} = \{Y_1(1)+Y_2(2), Y_2(2)\}$ at s=2, $\{Y_1(1)+Y_2(2), Y_1(2)+Y_2(3), Y_1(3)\}$ at s=3 or more generally $\{Y_1(1)+Y_2(2), Y_1(2)+Y_2(3),...., Y_1(s-1)+Y_2(s), Y_1(s)\}$ at s. A numerical example is given in Section 4.6. ■

The sufficient statistic at decision time s is SuffRq(s,t) t=1,...s, where SuffRq(s,t)$=\left(Y_1(t)+Y_2(t+q)\right)/2$ for t≤s-q and SuffRq(s,t)$=Y_1(t)$ for t>s-q. In Example 1 we have {SuffR1(1,t)}= {{Y1(1)} at s=1.

At s=2 we have {SuffR1(2,t)}= $\{[Y_1(1)+Y_2(2)]/2, Y_2(2)\}$.

At s=3 we have {SuffR1(3,t)}={{Y1(1)+Y2(2)]/2, [Y1(2)+Y2(3)]/2, Y1(3)}. More generally we have {SuffRp1(p,t)}={[$Y_1(1)+Y_2(2)$]/2, ...[$Y_1(2)+Y_2(3)$]/2,...., [Y_1(s-1) +Y_2(s)]/2, Y_1(s) }.

4. Surveillance methods for multivariate outbreak detection

In this section we will first describe the univariate outbreak detection method, OutbreakP, suggested by Frisén and Andersson (2009). Then, we will review common approaches to adapting univariate surveillance to multivariate surveillance and show how OutbreakP can be adapted by these approaches. After that, we will derive a joint multivariate method based on the sufficiency principle. Finally, we will give the maximum likelihood estimator of the parameters and a generalized likelihood ratio method for outbreak detection.

4.1. Univariate outbreak detection

For the outbreak detection situation, one way to specify the in-control state versus the outbreak is to use a parametric model of the outbreak curve. This requires extensive modeling as in for example Held et al. (2006). Here we will use a non-parametric univariate method as a base for the suggested adaption to a multivariate situation. When seasonal or other components are important, it might be useful to apply the non-parametric method to the residuals of a more complex model.

For the case of unknown parameters, generalized likelihood ratios (GLR) can be used by substituting the parameters with the maximum likelihood estimates. Lai (1995) suggested that in the CUSUM method, GLR should be used to handle unknown parameters after the change. This approach was also used by Höhle and Paul (2008) for Poisson and negative binomial distribution at surveillance of infectious diseases. In Frisén and Andersson (2009) a method for outbreak detection was suggested. The method utilized the GLR approach by using the maximum likelihood estimators under the monotonicity restrictions in Section 2.1, as derived in Frisén et al. (2010a) for the exponential family. The method was derived for the normal and Poisson distributions and was named the OutbreakP method for the Poisson distribution. Here, we will only consider the Poisson distribution, which is suitable for the application in Section 6. The method is semi-parametric since the distribution is parametric, but the regression is non-parametric since the only restriction on the regression is by monotonicity. A user-friendly

computer program can be downloaded at `www.statistics.gu.se/surveillance`. The method is also available in the `R package Surveillance`, described in Höhle (2010) and available on CRAN, and the open `JAVA package CASE` described in Cakici et al. (2010).

For the univariate surveillance of the influenza incidence in Sweden as a whole, the OutbreakP method was evaluated by Frisén and Andersson (2009) and Frisén, et al. (2009). We will now adapt this method for a multivariate situation.

4.2. General approaches to adapting univariate surveillance to multivariate surveillance

There are several approaches to multivariate surveillance. The most commonly used approach is the reduction to one scalar statistic, such as the sum for each time. This will be described in Section 4.3. Another approach is to use several univariate systems in parallel, one for each process. An intermediate approach is vector accumulation, for example MEWMA suggested by Lowry et al. (1992). When the multivariate distribution is available, as in e.g. Paul (2008), this might be used as a base for a surveillance method. An important situation treated by e.g. Tartakovsky and Veeravalli (2008) is where change in only one location can be expected and the identification of the correct one is crucial. General reviews on multivariate surveillance methods can be found for example in Basseville and Nikiforov (1993), Sonesson and Frisén (2005), Bersimis et al. (2007) and Frisén (2010).

4.3. Reduction to one scalar statistic for each time

Dimension reduction is always a reasonable choice in multivariate problems provided that it does not reduce important information. The most far-going reduction is the reduction to a scalar for each time. This is the most common way to handle multivariate surveillance. The observations at each time point consist of a vector, and we can first transform the vector from the current time point into a scalar statistic, which we then accumulate over time. In Sullivan and Jones (2002) this is referred to as "scalar accumulation". One natural reduction when dealing with multivariate normal variables is to use the Hotelling T^2 statistic suggested by Hotelling (1947). The Hotelling T^2 statistic is defined as $T^2(t) = (\mathbf{Y}(t) - \boldsymbol{\mu}_0(t))^T \mathbf{S}_{\mathbf{Y}(t)}^{-1} (\mathbf{Y}(t) - \boldsymbol{\mu}_0(t))$, where $\mathbf{S}_{\mathbf{Y}(t)}$ is the sample covariance matrix. Originally, the Hotelling T^2 statistic was used in a Shewhart approach, and this is sometimes referred to as the Hotelling T^2 control chart.

One example of scalar accumulation is when, for each time point, a statistic representing the important aspects of the spatial pattern is constructed from a purely spatial analysis. This statistic is then used in a surveillance method. The reduction to a univariate variable can be followed by univariate monitoring of any kind. In Rogerson (1997) and Rogerson (2001), different statistics measuring clustering were used for each time, and the information was accumulated by the univariate CUSUM method. In Zhou and Lawson (2008), the spatial pattern was characterized by a Bayesian model for each time, and the statistic was then monitored by the EWMA method.

For the influenza incidence, a natural reduction is the sum, even though information on different parts of the country is available. Using the sum means that no regional information is used. Instead, the surveillance is based on total data for the country as a whole, as in Frisén and Andersson (2009). However, other reductions may be more efficient, as is seen in Section 3. In our evaluations in Section 5, the reduction to a scalar is included.

4.4. Parallel outbreak detection

To illustrate a frequently used approach to multivariate surveillance, we will include a parallel system in our evaluations. By the parallel approach, each process is monitored separately and an overall alarm is called if some condition is fulfilled. The most common condition is that one of the systems calls an alarm. We will use this condition when the univariate OutbreakP method is applied to each process. An overall alarm is called the first time that any of the processes gives an alarm. The method is called OutbreakPParallel. Results for this method, as compared to others, are given in Section 5.3.

4.5. Outbreak surveillance based on sufficient reduction and known parameters

The likelihood ratio of an outbreak versus no outbreak with onsets of the outbreaks at τ_1, τ_2,... τ_p, is

$$L(s,t_1,...,t_p)=\frac{f(\mathbf{Y}^s\,|\tau_1=t_1,...,\tau_p=t_p)}{f(\mathbf{Y}^s\,|\tau_1>s,...,\tau_p>s)}$$

For known time lags *(q_1=0,q_2, q_3,... q_p),* this can be written

$$\mathrm{L(s,t_1)}=\frac{f(\mathbf{Y}^s\mid\tau_1=t_1)}{f(\mathbf{Y}^s\,|\tau_1>s)}$$

For detection of an outbreak as defined in Section 2 L(s,1) is the relevant statistic, see Frisén and Andersson (2009). For the Poisson distribution and known values of the parameters of the regressions, we have that

$$\mathrm{L(s,1)}=\prod_{i=1}^{p}\prod_{t=1+q_i}^{s}\exp(\lambda_0-\lambda_{t-q_i})\left(\frac{\lambda_{t-q_i}}{\lambda_0}\right)^{Y_i(t)}=\prod_{t=1}^{s}\mathrm{e}^{|I_t|(\lambda_0-\lambda_t)}\left(\frac{\lambda_t}{\lambda_0}\right)^{\sum_{i\in I_t}Y_i(t+q_i)},$$

where $I_t=\{i:q_i\le s-t,1\le i\le p\}$.

For two processes we have

$$\mathrm{L(s,1)}=\prod_{t=1}^{s-q}e^{2(\lambda_0-\lambda_t)}\left(\frac{\lambda_t}{\lambda_0}\right)^{Y_1(t)+Y_2(t+q)}\prod_{t=s-q+1}^{s}e^{\lambda_0-\lambda_t}\left(\frac{\lambda_t}{\lambda_0}\right)^{Y_1(t)}.$$

In Section 4.7 we will use the generalized maximum likelihood and substitute the unknown parameters with their maximum likelihood estimators derived in Section 4.6.

4.6. Maximum likelihood estimation of the multivariate outbreak regression

If the distribution of the processes is not fully specified, the approach of the generalized likelihood ratio can be used. Hence, we need estimates for the likelihood ratio in Section 4.5, both for the situation with an outbreak and for the situation with no outbreak. When we have no outbreak, and thus all observations are independent and identically distributed, the maximum likelihood estimator of λ_0 is the average of all observations. We have

$$\hat{\lambda}_0=\sum_{t=1}^{s}\sum_{i=1}^{p}y_i(t)/sp\,.$$

In the outbreak situation, we have the monotonicity restriction described in Section 2. A useful technique to find least squares estimates, which here are maximum likelihood estimates, is the Pool Adjacent Violator Algorithm, PAVA, described for example by Robertson et al. (1988).

Theorem 2: For the multivariate outbreak regression in Section 2.2 with processes which all belong to the one-parameter exponential family and which are independent and identically distributed, conditional on the change points and time lags (independent over time as well as across processes), the maximum likelihood estimators of λ_t, for the increasing phase are obtained by the PAVA algorithm with weights proportional to the number, $|I_t|$, of processes used for the specific component of the sufficient statistic.

Proof: In order to obtain the maximum likelihood estimators of the expected values λ_t for $\tau_1=1$, we utilize the assumption $\lambda_0 \leq \lambda_1 \ldots \leq \lambda_s$. Frisén, et al. (2010a) demonstrated that in the univariate case, the maximum likelihood estimators of the expected values λ_t of the outbreak regression can be obtained by the PAVA algorithm. For p processes, with known lags $(q_1=0, q_2, q_3, \ldots q_p)$, any observation of $Y_i(t)$ such that $t<\tau_i$ is an observation with the expected value λ_0. In the same way, any observation of $Y_i(t)$ such that $\tau_i = t$ has the expected value λ_1 and so on until the last observations of $Y_1(s)$ and any other $Y_i(s)$ such that $\tau_i = \tau_1$, which are observations with the expected value λ_s. Thus, the number of observations, $|I_t|$, with expectation λ_t depends on t and $(q_2, q_3, \ldots q_p)$. It follows from results on isotonic regression, with different numbers of observations for different values of the independent variable (see for example Theorem 1.5.2 in Robertson, et al. (1988)), that the maximum likelihood estimators are obtained by the PAVA on the average of the observations of λ_t with weights proportional to the number of observations, $|I_t|$. ■

EXAMPLE 2

To illustrate how the sufficient reduction and PAVA are used, we give a simple example for two processes with lag q=1. SuffRq(s,t) is the sufficient reduction described in Section 3.2, where q indicates the lag between the two processes and s is the decision time. In Table 1, we illustrate how the sufficient statistic and the maximum likelihood estimators are calculated for a numerical example.

Table 1. For an example of observations on two processes we give the sufficient statistic SuffR1 for s=1, 2, 3, 4, 5 and the maximum likelihood estimate $\hat{\lambda}_t$ at s=5.

t	y_1	y_2	SuffR1(1,t)	SuffR1(2,t)	SuffR1(3,t)	SuffR1(4,t)	SuffR1(5,t)	$\hat{\lambda}_t$
1	4	2	4	2.5	2.5	2.5	2.5	2.25
2	3	1		3	2	2	2	2.25
3	3	1			3	3	3	2.25
4	1	3				1	1.5	2.25
5	6	2					6	6

The estimate of $\hat{\lambda}_0$ is the average of all observations. At s=5 we have $\hat{\lambda}_0$=2.6. To estimate $\hat{\lambda}_t$ at time s=5, we apply the PAVA to the sequence SuffR1(5,t), t=1,...5. We see that the first violation of the order restriction occurs at t=2, and hence we replace the observations by the weighted average, (2.5·2+2·2)/4=2.25. This does not violate the first observation, $Y_2(1)$, since

$2 \leq 2.25$. The observation at t=4 constitutes a violation, and hence we use $(3 \cdot 2 + 1.5 \cdot 2)/4 = 2.25$, which does not violate the order restriction of the previous observations. ■

4.7. Generalized likelihood ratio surveillance of multivariate outbreaks

We will use the generalized likelihood ratio, i.e. substitute parameter values by their maximum likelihood estimators, in our semi-parametric multivariate method.

By substituting the parameters of the outbreak regression in L(s,1) in Section 4.5 with the maximum likelihood estimators in Section 4.6, we get the alarm statistic of the multivariate OutbreakPSuffR method. Here P stands for the Poisson distribution while SuffR stands for the sufficient reduction in the multivariate case. The general method depends on the set of lags $(q_1=0, q_2, q_3, \ldots q_p)$ and has the alarm statistic

$$\prod_{i=1}^{p}\prod_{t=1+q_i}^{s}\exp(\hat{\lambda}_0-\hat{\lambda}_{t-q_i})\left(\frac{\hat{\lambda}_{t-q_i}}{\hat{\lambda}_0}\right)^{Y_i(t)}=\prod_{t=1}^{s}\mathrm{e}^{|I_t|(\hat{\lambda}_0-\hat{\lambda}_t)}\left(\frac{\hat{\lambda}_t}{\hat{\lambda}_0}\right)^{\sum_{i\in I_t}Y_i(t+q_i)}$$

where $I_t=\{i: q_i \leq s-t, 1 \leq i \leq p\}$. For two processes with time lag q, we use the notation OutbreakPSuffRq for the method and OutbreakP SuffRq(s) for the alarm statistic. For this case we have

$$\prod_{t=1}^{s-q}e^{2(\hat{\lambda}_0-\hat{\lambda}_t)}\left(\frac{\hat{\lambda}_t}{\hat{\lambda}_0}\right)^{Y_1(t)+Y_2(t+q)}\prod_{t=s-q+1}^{s}e^{\hat{\lambda}_0-\hat{\lambda}_t}\left(\frac{\hat{\lambda}_t}{\hat{\lambda}_0}\right)^{Y_1(t)}$$

In the case q=0 this simplifies to the univariate OutbreakP statistic described in Frisén and Andersson (2009) and Frisén, et al. (2009).

EXAMPLE 3. For the situation of Example 1 and 2, we have for s=5 the alarm statistic

$$\text{OutbreakPSuffR1(5)}=\prod_{t=1}^{4}e^{2(\hat{\lambda}_0-\hat{\lambda}_t)}\left(\frac{\hat{\lambda}_t}{\hat{\lambda}_0}\right)^{Y_1(t)+Y_2(t+q)}\prod_{t=5}^{5}e^{\hat{\lambda}_0-\hat{\lambda}_t}\left(\frac{\hat{\lambda}_t}{\hat{\lambda}_0}\right)^{Y_1(t)}=6.14\,.\blacksquare$$

5. Simulation study to determine the properties of the multivariate OutbreakP method

In a multivariate situation, some reduction of the dimensionality of data is often useful, but it is important that no information is lost. This could be achieved by the use of a sufficient statistic. If the outbreaks appear simultaneously for the different processes, then we have a univariate sufficient statistic with one change point. However, when the outbreaks appear at different times, the sufficient statistic has more than one change point in the distribution. Even though each component has one change point, the distribution of the sufficient statistic is not constant either for $t < \tau_i$ or for $t \geq \tau_i$. The proofs commonly used for minimax or expected delay optimality require that there is only one change between two distributions.

Since exact optimality cannot be expected, the properties of the OutbreakP method are presented by the results from a simulation study. In Section 6 the method will be evaluated by the application of the method to observed Swedish influenza data.

5.1. *Model for simulations*

We used a model that is relevant for the application to the influenza data described in Section 6. The model is based on the study by Andersson, et al. (2008a) on the seasonal influenza in Sweden. The Poisson distribution was used for the incidences. The suggested method is non-parametric with respect to the shape. However, to examine the properties of the method by a simulation study, we used a parametric model to generate data. For the total influenza incidence in Sweden, the level at the constant phase, λ_0, is set to $\lambda_0 = 1$, and the parameter $\lambda(t)$ of the Poisson distribution follows an exponential curve $\lambda(t) = \exp(\beta_0 + \beta_1(t-\tau+1)$ for the increasing phase. The parameters were estimated to $\beta_0 = -0.26$ and $\beta_1 = 0.826$ from Swedish influenza data from the season 03-04, which was not extreme in any sense but "typical".

For the multivariate case, we use a model with two processes resembling those of the influenza data in Section 6. We use the results by Schiöler (2010) on how the incidence develops for the Metropolitan, M, and Local, L, areas, respectively. We use $E[M(t)]=0.5$ for $t<\tau$ and $E[M(t)]= \exp\{\beta_0+\beta_1(t-\tau+1)\}$, and $E(L(t)= 0.5$ for $t<\tau$ and $E(L(t))= \exp\{\beta_0+\beta_1 (t-\tau+1+q)\}$). With parameters, $\beta_0 = -0.622$ and $\beta_1 = 0.826$.

5.2. *False alarms*

The most commonly used measure for false alarms is the in-control average run length, ARL^0, $E[t_A|\tau=\infty]$. This can be used also in a multivariate situation. A similar measure, which is more convenient to calculate, is the median run length, MRL^0. We used the same MRL^0 (780) in all comparisons in this paper. It was used also for the univariate OutbreakP method in Frisén and Andersson (2009). The technique chosen by Frisén and Sonesson (2006) was used to ensure that the alarm limit was determined with enough accuracy to make the error in the curves of delay less than the line width.

5.3. *Delay*

One measure of the detection ability is the average run length, given that the change occurs immediately ($\tau=1$). This is widely used in univariate surveillance and often named zero-state ARL or ARL^1. Zero-state ARL is the most commonly used evaluation measure also in the multivariate case. However, it is seldom explicitly defined. The definition implicit in most publications is $E[t_A| \tau_1= \tau_{2=\ldots} \tau_p =1]$. Here, it is assumed that all processes change at the same time. As seen in Section 3.1, a sufficient reduction to a univariate problem exists when all processes change at the same time. Zero-state ARL is thus questionable as a formal measure for comparing methods for genuinely multivariate problems. Instead, we will here use a measure which allows different change points.

The conditional expected delay $CED(\tau) = E[t_A - \tau \mid \tau \le t_A]$ can be generalized for multivariate surveillance to $CED(\tau_1, \tau_{2\ldots} \tau_p) = E[t_A - \tau_{\min} \mid \tau_{\min} \le t_A]$, see Frisén, et al. (2010b). For a given lag this depends on only one of the change points. Thus we can write $CED(\tau_{\min}) = E[t_A - \tau_{\min} \mid \tau_{\min} \le t_A]$. When we have lag=0, i.e. simultaneous outbreaks, this reduces to the univariate CED. In Figure 1, we can see that the OutbreakPParallel method has a worse delay than the OutbreakPSuffR0 method for simultaneous outbreaks. OutbreakPSuffR0 is based on SuffR0, which corresponds to the total incidence. In Figure 2 we can see that the delay for the parallel method is worse than that for the OutbreakPSuffR1 method based on SuffR1 when lag=1.

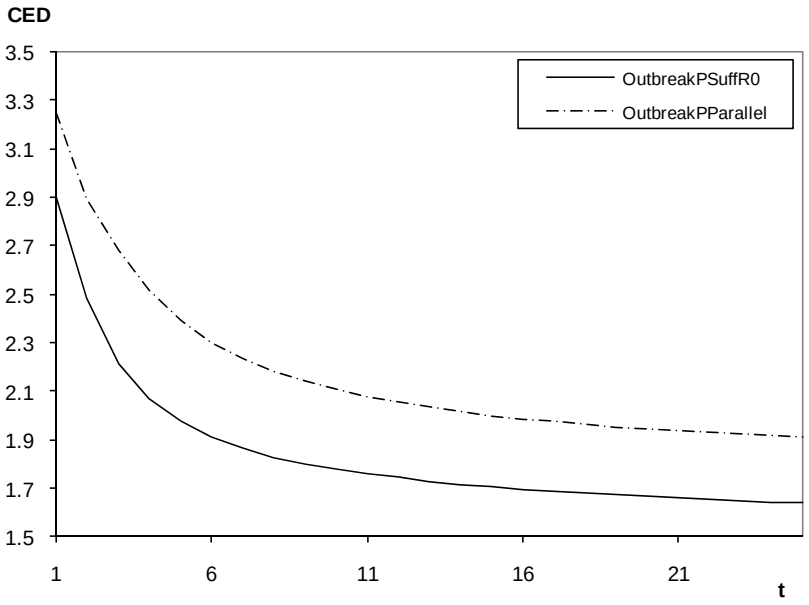

Fig. 1 The conditional expected delay for the OutbreakPParallel and OutbreakPSuffR0 methods for two processes with simultaneous onset of the outbreak (lag=0) as a function of τ_{min}=t.

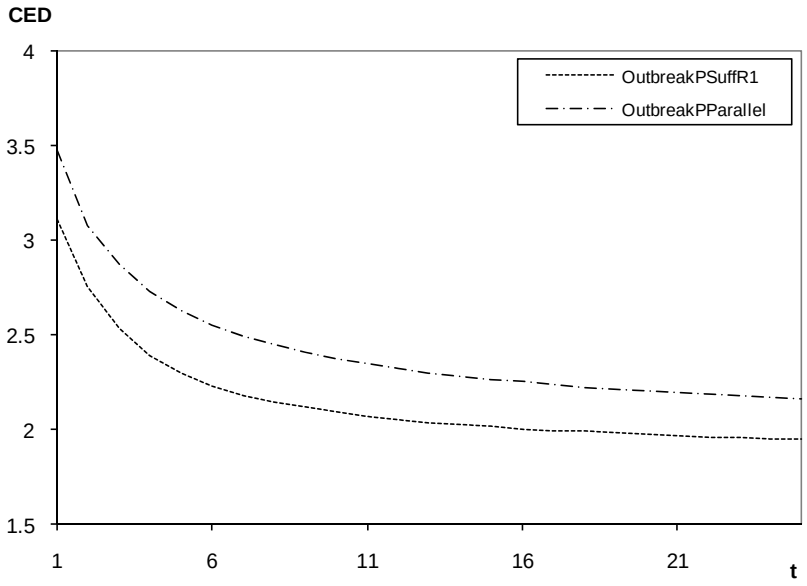

Fig. 2 The delay in detection of the outbreak for the OutbreakPParallel and OutbreakPSuffR1 methods for two processes with lag=1 as a function of τ_{min}=t.

5.4. Predictive value

If a method calls an alarm, it is important to know whether this alarm is a strong or weak indication of a change. The predictive value is a well-established measure in epidemiology. In surveillance, however, we need a variant that also incorporates time. The difference in surveillance, as compared to situations involving only one decision, is that we can get an alarm at any time point, and therefore we need a measure of the predictive value at each of them. In order to judge to what degree an alarm at time t_A can be trusted, it is necessary to consider the balance between the risk of false alarms, the detection ability and the probability of a change. If we have one change point τ and this is regarded as a random variable, this can be done by the probability of an outbreak, at an alarm, as suggested by Frisén (1992):

$$PV(t) = P(\tau \le t \mid t_A = t) = \frac{\sum_{i=1}^{t} P(t_A = t \,|\, \tau = i) P(\tau = i)}{\sum_{i=1}^{t} P(t_A = t \,|\, \tau = i) P(\tau = i) + P(t_A = t \,|\, \tau > t) P(\tau > t)}.$$

In a multivariate setting this was generalized by Frisén, et al. (2010b) to

$$PV(t) = P(\tau_{\min} \le t \mid t_A = t) = \frac{\sum_{i=1}^{t} \left(P(t_A = t \,|\, \tau_{\min} = i) P(\tau_{\min} = i) \right)}{\sum_{i=1}^{t} \left(P(t_A = t \,|\, \tau_{\min} = i) P(\tau_{\min} = i) \right) + P(t_A = t \,|\, \tau_{\min} > t) P(\tau_{\min} > t)}.$$

The predictive value depends on whether outbreaks appear frequently or rarely. Knowledge of the exact distribution of τ_{min} is seldom available, but we will nevertheless try to give a rough indicator. In the simulation study, τ_{min} was assumed to be geometrically distributed, i.e. $P(\tau_{\min} = i) = (1-\nu)^{i-1}\nu$. This may not give the closest fit of the onset times in Sweden, but in order to detect outbreaks which occur at unexpected times we did not want to include information on which week is the most common one for the onset. The level of intensity was roughly estimated from all available historical data on seasonal influenza to be $\nu = 0.1$. With this intensity the PV is above 0.99, and for a lower intensity, $\nu = 0.01$, which weakens the PV, it is above 0.95. The method and alarm limit used in the simulation study were considered potentially useful for practical application since the predictive value was high.

5.5. Robustness

Some models and assumptions are needed in order to efficiently make inferences from data. Hence, it is important to chose assumptions which are suitable for the application. Here we will concentrate on robustness related to a possible time lag. First we will describe the effect of using the method but with a wrong lag, then we will describe the consequences of different population sizes of different regions.

The lag between the outbreaks is seldom exactly known. We examined the effect of using the sufficient statistic for lag=1 when in fact lag=2, and vice versa. In Figure 3, we have simulated influenza outbreaks where the true lag is 1. We can see that when we used the method

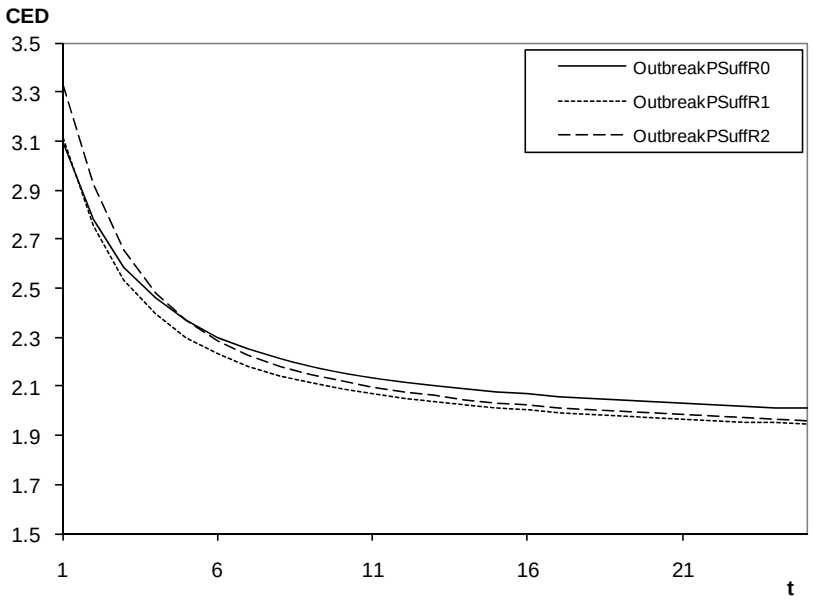

Fig. 3 The delay, as a function of τ_{min}=t, for outbreak detection by OutbreakPSuffR0, OutbreakPSuffR1 and OutbreakPSuffR2 when the true lag is 1.

OutbreakPSuffR1, which is based on the true lag, we got the best results. When we used the method for lag=2 or lag=0, the results were slightly worse. In Figure 4, we have simulated outbreaks with the true lag 2. When we used the outbreak detection method based on the true lag we got the best results, except for a very minor advantage for SuffR1 at τ=1 and 2. In this complex situation, the method based on the sufficient statistic is not always exactly optimal, but it usually works very well. When we used the statistic for lag=1 the results were similar to those for the true lag. However, when the lag was two steps away from the true one and we used the sufficient statistic for lag=0, while the true lag was 2, we got clearly worse results. The conclusion is that an approximate lag may work well, provided that it is not too far away from the true one.

In the simulation model used above, we assumed equal distributions given the possibly different times of onset. In practice, however, the two processes may be based on different population sizes or otherwise have different parameters. If the difference is large, this should be handled by adjustment of the weights and the alarm limit. The ratio in size between the two areas analyzed in Section 6 is approximately 1.17, and a suitable simulation model for this case was derived in Schiöler (2010). We examined what would happen if no adjustments were made and the same weights and alarm limit were used, as if the population sizes were the same. The OutbreakPSuffR methods performed slightly worse if different population sizes were used. However, the predictive value of an alarm was still greater than 0.99 for the intensity 0.10. The conclusion is that the predictive value did not change much and that the interpretation of the results would not be dramatically changed.

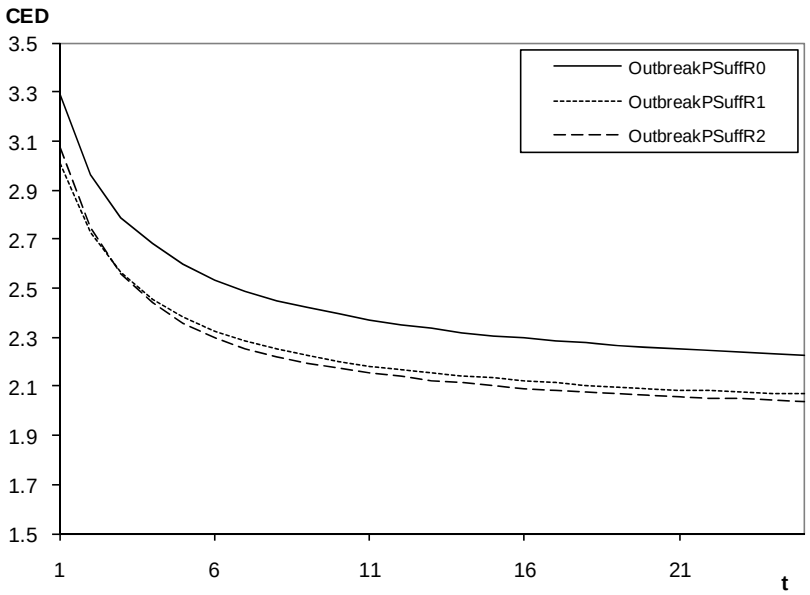

Fig. 4 The delay, as a function of τ_{min}=t, for outbreak detection by OutbreakPSuffR0, OutbreakPSuffR1 and OutbreakPSuffR2 when the true lag is 2.

6. Application of the multivariate OutbreakP method to Swedish regional influenza data

There are several national and international institutes that collect data on epidemic diseases, for example the European Centre for Disease Prevention and Control in Europe and the Centers for Disease Control and Prevention in the US. The monitoring of influenza in Sweden is mostly based on reports from all Swedish laboratories providing laboratory diagnoses of influenza (LDI). We will use these LDI data to illustrate the proposed method. In Sweden, data of infectious diseases are collected by the Swedish Institute for Infectious Disease Control, SMI. Andersson, et al. (2008a) and Andersson, et al. (2008b) give descriptions of the collection of these data. Here we use the laboratory-confirmed incidences of influenza type A or B. For some purposes, it may be of interest to monitor each location separately. However, the aim here is to get an alarm when the influenza epidemic has reached any part of Sweden. This means that the aim is to detect the first outbreak.

6.1. The spreading pattern of influenza in Sweden

The spatial pattern of how a disease spreads between regions is important. Spatial clustering of adverse health events is discussed for example by Kulldorff (2001), Rogerson (2001), Lawson and Rodeiro (2004), Marshall et al. (2007) and Sonesson (2007). However, in some situations, such as in the case of influenza in Sweden, the outbreak pattern is not characterized by clustering.

The spread of epidemic diseases, such as influenza, often follows geographical patterns. Schiöler (2010) searched for geographical patterns in the spread of influenza in Sweden (for example a pattern from south to north or from west to east). No such pattern was found. Instead it was found that influenza epidemics tend to start in the larger cities and then spread

to the smaller ones. Data from areas classified as Metropolitan areas generally showed an earlier outbreak than those from the Locality areas. The Metropolitan areas have major international airports nearby (Arlanda, Landvetter, Umeå and Kastrup), and commuting to other countries is common. This is a plausible explanation for the early start of the influenza season in these areas. This is also in accordance with the results of Crepey and Barthelemy (2007), who investigated the relation between travelling and influenza in the US and in France and found a stable impact.

The time difference in the onset of the influenza outbreak was about one week. This information will be used to increase the efficiency of our surveillance system.

6.2. Outbreak detection of influenza in Sweden

Based on the results on sufficiency in Section 3, the maximum likelihood estimation in Section 4.6, the generalized likelihood ratio in Section 4.7 and the choice of alarm limit in Section 5 to give MRL^0=780 and a predictive value greater than 90 %, we applied the OutbreakPSuffR1 to 11 seasons of influenza.

Figure 5 shows the results for the season 06-07. By accumulating the information by the OutbreakPSuffR1 alarm statistic, the outbreak is more clearly seen than when by the statistic based on the total number of cases in Sweden.

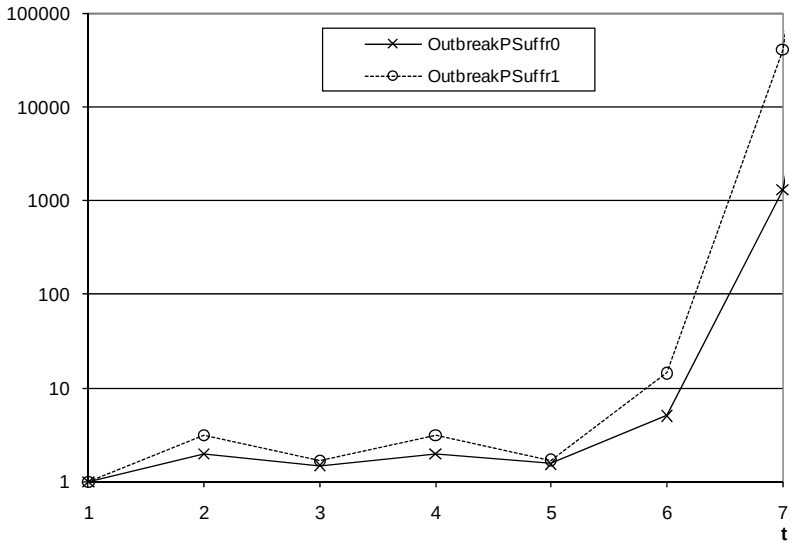

Fig. 5 The alarm statistic of the OutbreakPSuffR1method compared to that of OutbreakPSuffR0 up to the week of alarm during the season 06-07.

The situation varies from year to year. In Table 2, the week of the alarm is given for OutbreakPSuffR0 and OutbreakPSuffR1 for all years with available data. The alarm limits were chosen by way of the simulation study in Section 5 to have the same false alarm property with MRL^0=780. The OutbreakP based on SuffR1 gives an alarm the same week or earlier compared to OutbreakP based on the SuffR0, the total. As can be seen from the table, the alarm is given at the same time for eight seasons and earlier for three seasons for OutbreakP based on SuffR1 as compared to SuffR0. Note that the last season differs from the earlier ones due to the new H1N1 influenza. The incidences (of influenza type A or B) were very low this season and highly dominated by the metropolitan areas. This explains why there was an alarm

of an outbreak by the OutbreakSuffR1 method, which utilizes information on the metropolitan areas, but not by OutbreakSuffR0, which uses only the total for the country as a whole.

Table 2. Results for 11 influenza seasons in Sweden. The week of alarm is given for the methods based on the SuffR0 and SuffR1, respectively. The last column shows which method gave the first indication of an outbreak.

Season	SuffR0	SuffR1	First
99_00	49	49	Same
00_01	52	52	Same
01_02	2	2	Same
02_03	1	1	Same
03_04	46	46	Same
04_05	50	48	SuffR1
05_06	1	1	Same
06_07	47	46	SuffR1
07_08	51	51	Same
08_09	48	48	Same
09_10	No alarm	24	SuffR1

7. Discussion

In recent years, there have been several events that highlight the importance of outbreak detection. The outbreaks of new kinds of influenza (SARS, avian and H1N1) are such recent examples.

Several different definitions of an outbreak are used, explicitly or implicitly, in literature. Three commonly used approaches to outbreak detection are: i) the detection of an increasing incidence, ii) the detection of an incidence that is higher than expected, based on the information available up to that point and iii) the detection of a spatial clustering of cases which results in a higher incidence in an area than in its surroundings. The choice of method and evaluation procedure depends on which definition is used. Therefore, it is important to state the aim explicitly. Different methods may be optimal under different conditions, which means that the methods can often be seen as complements to each other.

The semi-parametric method used here detects outbreaks defined as a monotonic increase following the constant level before the onset of the outbreak. Such outbreaks are of interest in connection with several diseases and syndromes. Often, the information about the baseline is limited. Errors in the estimation of the baseline can have serious effect, as demonstrated for example by Frisén and Andersson (2009). Also, there may be seasonal effects with the same periodicity as the disease as well as large variation between years, thus making it hard to state the expected incidence. Therefore, it can be of value to have access to a method, which does not require knowledge about the baseline but is focused on the increasing incidence at an outbreak. A semi-parametric maximum likelihood ratio surveillance method was derived in Frisén and Andersson (2009) for the regular exponential family and applied and compared in Frisén, et al. (2009). The likelihood principle makes it possible to include knowledge on the probability of an outbreak depending on the season. However, here we chose a non-informative approach, since it may be valuable to detect outbreaks that occur at unexpected times.

When data from different sources are available, multivariate surveillance should be applied. This is the case for detection of influenza outbreaks on the basis of data from different regions. The two simplest approaches of multivariate surveillance are the reduction to a suitable univariate statistic and parallel surveillance with due concern to the multiplicity. We included these approaches in our evaluations by simulations. We also suggested a joint generalized likelihood ratio method based on maximum likelihood under multivariate monotonicity restrictions. The properties depend heavily on the relation between the times of onset in the different processes.

The relation between different processes is important in multivariate surveillance, as demonstrated by e.g. Frisén, et al. (2010b). The method that is optimal for simultaneous changes is not efficient at a time lag. The exact relation between the onset on different location is seldom exactly known. However, there can be some information as demonstrated in e.g. Schiöler (2010) where it was found that the influenza outbreak in Sweden in general started a week earlier in major cities than the rest of the country. In the application to the Swedish influenza data it was demonstrated that the performance of the surveillance was improved by utilizing this knowledge. The simulation study demonstrated that the even if the true time lag is only approximately known it can be an improvement to use it in the method.

Most theory of statistical surveillance is based on a change between two distributions – one for the times before the change point and another for the times after it. For simultaneous changes, we demonstrated that the sufficient statistic has one change point and that the suggested method is optimal. However, when changes occur at different times we can have several changes in the multivariate distribution. Thus, we cannot expect optimality. Here, we demonstrated that the suggested method gave good results both in the simulation study and when applied to spatial information on influenza in Sweden. We used a simulation model mimicking the behavior of Swedish influenza data, based on the results of Andersson, et al. (2008a), where a discussion on data quality problems was included. When evaluating methods for on-line monitoring it is important to use measures that incorporate the time issue, i.e. the fact that there are repeated decisions, not just one decision as in hypothesis testing. Here, we used evaluation measures by Frisén, et al. (2010b), which are better suited for multivariate on-line surveillance than the conventional ones.

The primary motive for this paper was the need for spatial surveillance of influenza outbreaks in Sweden. The suggested method may also be useful for other applications. The case of proxy data for influenza was discussed in Section 2.2. The detection of a change from a constant level to a monotonic trend is of special interest in connection with outbreaks of epidemic diseases. However, it may be useful also in other areas. For example, Schiöler and Frisén (2008) discussed the application of the outbreak method for detecting a decline in the results of financial managers.

Acknowledgements. Eva Andersson and Kjell Pettersson have given constructive comments. The data were made available to us by the Swedish Institute for Infectious Disease Control, and we are grateful for discussions about the aims and the data quality. The work was supported by the Swedish Civil Contingencies Agency (grant 0314/206).

References

ANDERSSON, E., BOCK, D. and FRISÉN, M. (2008a). Modeling influenza incidence for the purpose of on-line monitoring. *Statistical Methods in Medical Research,* **17** 421-438.

ANDERSSON, E., KUHLMANN-BERENZON, S., LINDE, A., SCHIÖLER, L., RUBINOVA, S. and FRISÉN, M. (2008b). Predictions by early indicators of the time and height of yearly influenza outbreaks in Sweden. *Scandinavian Journal of Public Health,* **36** 475-482.

BASSEVILLE, M. and NIKIFOROV, I. (1993). *Detection of abrupt changes- Theory and application*. Prentice Hall, Englewood Cliffs.

BERSIMIS, S., PSARAKIS, S. and PANARETOS, J. (2007). Multivariate Statistical Process Control Charts: An Overview. *Quality and Reliability Engineering International,* **23** 517-543.

CAKICI, B., HEBING, K., GRÜNEWALD, M., SARETOK, P. and HULTH, A. (2010). CASE –a framework for computer supported outbreak detection. *BMC Medical Informatics and Decision Making,* **10**.

CREPEY, P. and BARTHELEMY, M. (2007). Detecting Robust Patterns in the Spread of Epidemics: A Case Study of Influenza in the United States and France. *American Journal of Epidemiology,* **166** 1244-1251.

FRISÉN, M. (1992). Evaluations of methods for statistical surveillance. *Statistics in Medicine,* **11** 1489-1502.

FRISÉN, M. (2003). Statistical surveillance. Optimality and methods. *International Statistical Review,* **71** 403-434.

FRISÉN, M. (2009). Optimal sequential surveillance for finance, public health and other areas. Editor's special invited paper. *Sequential Analysis,* **28** 310-337, discussion 338-393.

FRISÉN, M. (2010). Principles for Multivariate Surveillance. In *Frontiers in Statistical Quality Control 9 (*H.-J. LENZ, P.-T. WILRICH and W. SCHMID, eds.) 133-144. Physica-Verlag, Heidelberg.

FRISÉN, M. and ANDERSSON, E. (2009). Semiparametric surveillance of monotonic changes. *Sequential Analysis,* **28** 434-454.

FRISÉN, M., ANDERSSON, E. and PETTERSSON, K. (2010a). Semiparametric estimation of outbreak regression. *Statistics: A Journal of Theoretical and Applied Statistics,* **44** 107 - 117.

FRISÉN, M., ANDERSSON, E. and SCHIÖLER, L. (2009). Robust outbreak surveillance of epidemics in Sweden. *Statistics in Medicine,* **28** 476-493.

FRISÉN, M., ANDERSSON, E. and SCHIÖLER, L. (2010b). Evaluation of Multivariate Surveillance. *Journal of Applied Statistics* to appear.

FRISÉN, M., ANDERSSON, E. and SCHIÖLER, L. (2010c). Sufficient reduction in multivariate surveillance. *Communications in Statistics-Theory and Methods* to appear.

FRISÉN, M. and SONESSON, C. (2006). Optimal surveillance based on exponentially weighted moving averages. *Sequential Analysis,* **25** 379-403.

GINSBERG, J., MOHEBBI, M. H., PATEL, R. S., BRAMMER, L., SMOLINSKI, M. S. and BRILLIANT, L. (2009). Detecting influenza epidemics using search engine query data. *Nature,* **457** 1012-1014.

HELD, L., HOFMAN, M., HÖHLE, M. and SCHMID, V. (2006). A two-component model for counts of infectious diseases. *Biostatistics,* **7** 422-437.

HOTELLING, H. (1947). Multivariate Quality Control. In *Techniques of statistical analysis (*C. EISENHART, M. W. HASTAY and W. A. WALLIS, eds.) 111-184. McGraw-Hill, New York.

HULTH, A., RYDEVIK, G. and LINDE, A. (2009). Web Queries as a Source for Syndromic Surveillance. *PLoS ONE,* **4** e4378.

HÖHLE, M. (2010). Aberration Detection in R Illustrated by Danish Mortality Monitoring. In *Biosurveillance (*T. KASS-HOUT and X. ZHANG, eds.) CRC Press.

HÖHLE, M. and PAUL, M. (2008). Count data regression charts for the monitoring of surveillance time series. *Computational Statistics & Data Analysis,* **52** 4357-4368.

JONER JR., M. D., WOODALL, W. H., REYNOLDS JR., M. R. and FRICKER, R. D. (2008). A One-sided MEWMA Chart for Health Surveillance. *Quality and Reliability Engineering International,* **24** 503-518.

JÄRPE, E. (2000). On univariate and spatial surveillance. Ph.D Thesis. Göteborg University, Göteborg.

KASS-HOUT, T. and ZHANG, X. (Eds.). (2010). *Biosurveillance: A Health Protection Priority*: CRC Press.

KULLDORFF, M. (2001). Prospective time periodic geographical disease surveillance using a scan statistic. *Journal of the Royal Statistical Society A,* **164** 61-72.

LAI, T. L. (1995). Sequential Changepoint Detection in Quality-Control and Dynamical Systems. *Journal of the Royal Statistical Society B,* **57** 613-658.

LAWSON, A. and RODEIRO, C. (2004). Developements in general and syndromic surveillance for small area health data. *Journal of Applied Statistics,* **31** 397-406.

LAWSON, A. B. and KLEINMAN, K. (Eds.). (2005). *Spatial and Syndromic Surveillance for Public Health*. New York: Wiley.

LOWRY, C. A., WOODALL, W. H., CHAMP, C. W. and RIGDON, S. E. (1992). A multivariate exponentially weighted moving average control chart. *Technometrics,* **34** 46-53.

MARSHALL, J. B., SPITZNER, D. J. and WOODALL, W. H. (2007). Use of the local Knox statistic for the prospective monitoring of disease occurrences in space and time. *Statistics in Medicine,* **26** 1579-1593.

PAUL, M., L. HELD, AND A. M. TOSCHKE. (2008). Multivariate modelling of infectious disease surveillance data. *Statistics in Medicine,* **27** 6250-6267.
ROBERTSON, T., WRIGHT, F. T. and DYKSTRA, R. L. (1988). *Order Restricted Statistical Inference*. Wiley, Chichester.
ROGERSON, P. A. (1997). Surveillance systems for monitoring the development of spatial patterns. *Statistics in Medicine,* **16** 2081-2093.
ROGERSON, P. A. (2001). Monitoring point patterns for the development of space-time clusters. *Journal of the Royal Statistical Society A,* **164** 87-96.
RYAN, T. P. (2000). Statistical methods for quality improvement. Wiley, New York.
SCHIÖLER, L. (2010). *Modelling the spatial patterns of influenza incidence in Sweden* (No. 2010:1). Gothenburg: Statistical Research Unit, Department of Economics, University of Gothenburg, Sweden.
SCHIÖLER, L. and FRISÉN, M. (2008). *On statistical surveillance of the performance of fund managers* (No. 2008:4): Statistical Research Unit, Department of Economics, University of Gothenburg, Sweden.
SHMUELI, G. and BURKOM, H. S. (2010). Statistical Challenges Facing Early Outbreak Detection in Biosurveillance. *Technometrics,* **52** 39-51.
SONESSON, C. (2007). A CUSUM framework for detection of space-time disease clusters using scan statistics. *Statistics in Medicine,* **26** 4770-4789.
SONESSON, C. and BOCK, D. (2003). A review and discussion of prospective statistical surveillance in public health. *Journal of the Royal Statistical Society A,* **166** 5-21.
SONESSON, C. and FRISÉN, M. (2005). Multivariate surveillance. In *Spatial surveillance for public health (*A. LAWSON and K. KLEINMAN, eds.) 169-186. Wiley, New York.
SULLIVAN, J. H. and JONES, L. A. (2002). A self-starting control chart for multivariate individual observations. *Technometrics,* **44** 24-33.
TARTAKOVSKY, A. G. and VEERAVALLI, V. V. (2008). Asymptotically Optimal Quickest Change Detection in Distributed Sensor Systems. *Sequential Analysis,* **27** 441 - 475.
WESSMAN, P. (1998). Some Principles for surveillance adopted for multivariate processes with a common change point. *Communications in Statistics - Theory and Methods,* **27** 1143-1161.
WOODALL, W. H. (2006). The Use of Control Charts in Health-Care Monitoring and Public-Health Surveillance. *Journal of Quality Technology,* **38** 89-134.
WOODALL, W. H. and MONTGOMERY, D. C. (1999). Research Issues and Ideas in Statistical Process Control. *Journal of Quality Technology,* **31** 376-386.
ZHOU, H. and LAWSON, A. B. (2008). EWMA smoothing and Bayesian spatial modeling for health surveillance. *Statistics in Medicine,* **27** 5907-5928.
ZHOU, Q., LUO, Y. and WANG, Z. (2010). A control chart based on likelihood ratio test for detecting patterned mean and variance shifts *Computational Statistics & Data Analysis,* **54** 1634-1645.

L. SCHIÖLER
M. FRISÉN
Statistical Research Unit
University of Gothenburg
SE 405 30 Gothenburg
Sweden
E-MAIL: linus.schioler@statistics.gu.se
marianne.frisen@statistics.gu.se

STATISTICAL RESEARCH UNIT, UNIVERSITY OF GOTHENBURG
Box 640, SE 405 30 Göteborg, Sweden

1. Hyrenius, H.: Statistical studies in the structure and recruitment of the clergy of the Church of Sweden. In Swedish with an English summary. Stockholm 1954. Pp. 63. (Out of print)

2. Hyrenius, H. & Zachrisson, U.: Studies of children in incomplete families I. Illegitimate children in Gothenburg 1928-54. In Swedish with an English summary. Stockholm 1955. Pp. 58. (Out of print)

3. Zachrisson, U. & Kjessler, Å.: Variations in prenatal mortality with social and economic factors and prenatal care. Göteborg 1956. Pp. 44. (Out of print)

4. Weibull, C.: The distribution of reciprocal choices in sociometric tests. Göteborg 1958. Pp. 16. (Out of print)

5. Hyrenius, H., Hugosson, K. & Gustafsson, SE.: Structure and changes of economy and population in the Östersund region. In Swedish with an English summary. Göteborg 1959. Pp. 142. (Out of print)

6. Zachrisson, U.: The distribution of "Student's t" in samples from individual non-normal populations. Göteborg 1959. Pp. 32.

7. Weibull, C.: Some aspects of statistical inference with applications to sample survey theory. Göteborg 1960. Pp. 87. (Out of print)

8. Hyrenius, H.: Flerspråkig demografisk ordbok. svensk volym. (Multilingual demographic dictionary. Swedish volume.) Göteborg 1961. Pp. 136. (See demographic Research Institute Reports No 1)

9. Zachrisson, U.: A study of truncation and censoring with economic subsidiary conditions. Göteborg 1962. Pp. 117.

10. Hyrenius, H. & Gustafsson, R.: Tables of normal and log-normal random deviates. Part I. Göteborg 1962. Pp. 103.

11. Hyrenius, H. & Gustafsson, R.: Tables of normal and log-normal random deviates. Part II. Göteborg 1962. Pp. 113.

12. Hyrenius, H., Adolfsson, I. et al (editors). Selected bibliography of non-normality. Göteborg 1964. Pp. 108.

13. Hyrenius, H.: Methods of valuating reindeer pastures – an economic study. In Swedish with an English summary. Göteborg 1964. Pp. 69.

14. Carlström, E. & Zachrisson, U.: On quality and costs in life testing. Göteborg 1965. Pp. 57.

15. Norlén, U.: Simulation Model Building. A statistical approach to modelling in the social sciences with the simulation method. Stockholm 1972. Pp. 172.

(cont.)

STATISTICAL RESEARCH UNIT, UNIVERSITY OF GOTHENBURG
Box 640, SE 405 30 Göteborg, Sweden

(Continuation)

16. Frisén, M.: Stochastic deviation from elliptical shape. Göteborg 1974. Pp. 63.

17. Klevmarken, A.: Ålders-, kvalifikations- och befordringstillägg. Göteborg 1980. Pp. 99.

18. Jonsson, R.: A model for miniature end-plate potentials based on the hypothesis of a release-regulating barrier. Göteborg 1981. Pp. 118.

19. Tómasson, H.: Prediction and estimation in ARMA models. Göteborg 1986. Pp. 118.

20. Johnsson, T.: On stepwise procedures for some multiple inference problems. Göteborg 1989. Pp. 145.

21. Svensson, E.: Analysis of systematic and random differences between paired ordinal categorical data. Göteborg 1993. Pp. 560.

22. Palaszewski, B.: On multiple test procedures for finding deviating parameters. Göteborg 1993. Pp. 120.

23. Dahlbom, U.: Estimation of regression functions with certain monotonicity and concavity/convexity restrictions. Göteborg 1994. Pp. 249.

24. Hussain, S.: On cointegration and persistence. Göteborg 1995. Pp. 202.

25. Wessman, P.: Studies on the surveillance of univariate and multivariate processes. Göteborg 1999. Pp. 100.

26. Järpe, E.: On Univariate and Spatial Surveillance. Göteborg 2000. Pp. 103.

27. Andersson, E.: On Turning Point Detection in Cyclical Processes. With applications to the monitoring of business cycles. Göteborg 2001. Pp. 171.

28. Holgersson, T.: On Assessing Distributional Properties of Multidimensional Variables. Göteborg 2003. Pp. 137.

29. Petzold, M.: Evaluation of Information in Longitudinal Data. Göteborg 2003. Pp. 112.

30. Sonesson, C.: On Statistical Surveillance. Issues of Optimality and Medical Applications. Göteborg 2003. Pp. 138.

31. Bock, D: Statistical surveillance. Optimal decisions times in economics. Göteborg 2004. Pp. 140.

ISBN 978-91-85169-54-2