# Aberration Detection in R Illustrated by Danish Mortality Monitoring

Michael Höhle

*Department of Statistics, Ludwig-Maximilians-Universität München, Munich, Germany*
*Munich Center of Health Sciences, Munich, Germany*

Anne Mazick

*Department of Epidemiology, Statens Serum Institute, Copenhagen, Denmark*

*The present text is a preprint of a book chapter to appear in T. Kass-Hout and X. Zhang (Eds.) Biosurveillance: A Health Protection Priority, CRC Press.*

## 1. Introduction

The objective of biosurveillance in this chapter is the detection of emerging incidence clusters in time of a health related event. Reviews on temporal surveillance can be found e.g. in Sonesson and Bock (2003), Bravata et al. (2004), Buckeridge et al. (2005) and Tennant et al. (2007). In recent years a pleasant development has been a synthesis of surveillance methods with methods from statistical process control, see e.g. Woodall (2006) for a survey.

One important aspect to ensure a transfer of methodological developments into practice is the availability of appropriate software implementations and their documentation. With the present chapter we want to introduce one such open-source software implementation into a public health context: the R package `surveillance`. In order to demonstrate functionality, we use Danish mortality data from the ongoing European monitoring of excess mortality for public health action (EuroMOMO) project (Anonymous, 2009).

The R system is a free software environment for statistical computing and graphics distributed under a GNU-style copyleft license and running under Unix, Windows and Mac (R Development Core Team, 2009). Several documents and books provide an introduction (Dalgaard, 2008; Venables et al., 2009; Muenchen, 2009). The add-on package `surveillance` offers functionality for the visualization, monitoring and simulation of count data time series in R for public health surveillance and biosurveillance. It provides an implementation of different aberration detection algorithms for epidemiologists and an infrastructure for developers of new algorithms. The package is freely available under the GNU GPL license and obtainable from the Comprehensive R Archive Network (CRAN). To install the package from CRAN, the following call in R has to be performed once:

```
R> install.packages("surveillance")
```

After installation the package is loaded using:

```
R> library("surveillance")
```

Focus in the present chapter is on using the aberration detection algorithms in the package for univariate count data time series, but the package also contains example outbreak data from the German SurvStat@RKI database (Robert Koch Institute, 2009), functionality for

the simulation of outbreak data and the comparison of algorithms. Höhle (2007) provides basic information about the package, further information can be found at the package homepage located at `http://surveillance.r-forge.r-project.org/`. The present text introduces a number of new developments in the package, e.g. using the S4 `sts` class for gathering data and methods and using likelihood ratio based cumulative sum (CUSUM) algorithms.

At the time of writing, only few other R packages exist aimed at helping epidemiologists in their outbreak detection and outbreak investigation. Retrospective cluster detection is available for example in the `DCluster` package (Lopez-Quílez, 2005) and visualization of outbreak data can be performed by `epitools` (Aragon, 2008). Retrospective and – to some extent – prospective investigations of structural changes in time series can also be performed by the package `strucchange` (Zeileis et al., 2002), which, however, aims more at the econometrics community.

### *1.1. The EuroMOMO project*

The project "European monitoring of excess mortality for public health action" (EuroMOMO) is a three year project representing a network of 23 partners from 21 countries in the European region. The project is co-funded by the European Commission and coordinated by Statens Serum Institut, Denmark (Mazick, 2007; Anonymous, 2009).

The aim of EuroMOMO is to develop and strengthen real-time monitoring of mortality across Europe; this will enhance the management of serious public health risks such as pandemic influenza, heat waves and cold snaps. EuroMOMO's general objective is to develop and operate a routine public health system that monitors all-cause mortality in order to detect and measure – in a timely manner – excess number of deaths related to influenza and other known or emerging public health threats across European countries. Main actions include the creation of an inventory of existing mortality monitoring systems in Europe; the definition of minimal requirements for a mortality monitoring system; retrospective analysis of mortality data; identification of an optimal common analytical approach and piloting of such a consensus system for mortality monitoring in several European countries.

Mortality monitoring is useful for early detection and monitoring of severe impacts of health threats and is as such an indicator-based surveillance system that provides important information within the framework of epidemic intelligence. The latter comprising the collection, collation, analysis and assessment of information from different sources to rapidly identify and respond to known and unknown public health threats (Kaiser et al., 2006). Vital statistics are accessible for all European countries. However, often these data are not readily available in a timely manner during health crises or for imminent health threats. On the other hand, decision makers will request up-to-date mortality data in case of the threat of epidemics or emergence of new diseases (e.g. pandemic influenza, AIDS or SARS). As these threats are not restricted by borders, not only a national but also a common European approach to detect and estimate the magnitude of deaths is required. This is especially important as the methodology of monitoring mortality is complex and there is a risk of European countries sharing incompatible information if different methodologies are used. However, in Europe, real-time monitoring of mortality is presently neither carried out uniformly nor in many European countries.

The main outcome of mortality monitoring is excess mortality, which can be defined as observed mortality in a given time period, e.g. a week, minus the expected mortality for that time period. Ongoing data analysis involving modeling the expected number of deaths

for a given geographical unit and for different population groups is needed, and there are several candidate models available that have been tested or are in use in a few European countries (Conti et al., 2005; Gergonne et al., 2005; Josseran et al., 2006; Nogueira et al., 2005; Simón et al., 2005). However, in order to compare estimates of excess deaths a common versatile statistical model is needed, and the key output of EuroMOMO is to provide a European consensus model for mortality monitoring which is applicable all over Europe and which is piloted and ready to implement.

The structure of this chapter is as follows. Section 2 provides an overview of the `surveillance` package, Section 3 constitutes the statistical framework of aberration detection, which is then considered in details for the Farrington method and the negative binomial CUSUM in Sections 4 and 5. After a description of the theory and syntax of invocation, Danish mortality data are used in each section to illustrate the methods. Section 6 concludes the chapter with a discussion.

## 2. Overview of the `surveillance` Package

The functionality in `surveillance` can be divided into two categories: Prospective changepoint (aka. aberration) detection algorithms for univariate time series of counts and retrospective modeling of possibly multivariate times series of counts.

Classical public health aberration detection algorithms for univariate time series found in `surveillance` are e.g. the function `cdc` implementing the approach described in Stroup et al. (1989) and the function `farrington` implementing the work of Farrington et al. (1996). More statistical process control oriented approaches can be found as functions `cusum` (Rossi et al., 1999), `rogerson` (Rogerson and Yamada, 2004) and `glrnb` (Höhle and Paul, 2008).

Retrospective time series modeling is available in: `algo.hmm`, implementing the hidden Markov model approach in Le Strat and Carrat (1999) and `algo.hhh`, implementing the branching process approach described in Held et al. (2005) and Paul et al. (2008). Furthermore, `algo.twins` contains an implementation of the two-component endemic and epidemic approach described in Held et al. (2006).

In what follows, focus will be on aberration detection methods. A prerequisite to their use is an understanding of the data structure and related access and visualization methods for the data.

### *2.1. Data structure and data input*

The S4 class `sts` (an abbreviation for surveillance time series) provides a data structure for handling the multivariate time series of counts of the form $\{y_{it}\,;\, i = 1, \ldots, m,\, t = 1, \ldots, n\}$. Here $n$ denotes the length of the time series and $m$ denotes the number of entities being monitored, e.g. geographical regions, hospitals or age groups. A slot `observed` of `sts` contains an $n \times m$ matrix representing the $y_{it}$ counts. The slot `start` denotes the origin of the time series given by a vector of length two containing the year and the epoch within that year. Furthermore, `freq` denotes the number of observations per year, e.g. 365 for daily data, 52 for weekly data and 12 for monthly data. An integer slot `epoch` denotes the time index $1 \leq t \leq n$ of each row in `observed`.

To import data into R and `surveillance`, one can use R's `read.table` or `read.csv` functions to read ASCII text or comma separated value files. A different option is to use the package `foreign` to import SAS, SPSS, Stata or dBase files or the `RODBC` database

interface to import from Microsoft Access/Excel or SQL databases. An `sts` object is then created from the resulting matrix of counts. We start the analysis of the Danish 1994-2008 mortality data by reading a CSV file (782 rows and 8 columns) containing the weekly number of all-cause mortality and use this to create an `sts` object.

```
R> momo.ts <- read.csv("mortality-dk.csv", header = TRUE, check.names = FALSE)
R> dates <- as.Date("1994-01-03") + 7 * 0:(nrow(momo.ts) - 1)
R> momo <- new("sts", epoch = as.numeric(dates), start = c(1994, 1),
+     freq = 52, observed = momo.ts, epochAsDate = TRUE)
```

The eight columns correspond to the eight age groups <1, 1-4, 5-14, 15-44, 45-64, 65-74, 75-84 and ≥85 years as defined by the EuroMOMO project to be a relevant age stratification. Deaths are registered by the day of death. A special feature of the EuroMOMO data is that weeks are handled as defined by the ISO 8601 standard (Anonymous, 2004). This standard defines week-numbering for a year to start at the first Monday of week 01 and to end at the last Sunday before the new ISO year. Here, week 01 of a year is the week with the year's first Thursday in it. As a consequence, a year consists of either 52 or 53 full weeks. Usually, one operates in `surveillance` with a fixed number of epochs per period, e.g. 52 weeks per year as given by the `freq` argument. But by specifically setting the `epoch` slot to a numeric representation of the corresponding Monday of each week and setting the `epochAsDate` attribute, we can use the `Date` class in R to easily handle this ISO week complication.

The resulting `sts` object `momo` can now be accessed and manipulated using standard matrix and data frame like access, e.g. `momo[1:10,"[0,1)"]` gives an `sts` object containing the first 10 weeks of the <1 age group and `dim(momo)` returns the dimension of the momo time series (i.e. $782 \times 8$). Other operations are the aggregation of the time series over several epochs or entities by the `aggregate` function or linking the multivariate times series to geographical regions of an ESRI shapefile. Plot functions provide visualization of the multivariate time series in time, space and space-time. In the subsequent analysis of the Danish mortality data we focus on the country aggregated time series stratified by age. Here, age stratification is used to differentiate between different mortality risk groups, and country level is used to ensure sufficiently large strata in a population of 5.5 millions. For larger EuroMOMO countries, a further stratification by geographical region might, however, be relevant. The following code illustrates various uses of the `plot` function for the `momo` object with corresponding output shown in Figs. 1 and 2.

```
R> plot(momo[year(momo) >= 2000, ], type = observed ~ time | unit)
R> plot(momo, ylab = "No. of deaths", type = observed ~ time)
R> plot(momo[, "[0,1)"], ylab = "No. of deaths")
```

In the above lines, the `type` argument controls the view on the multivariate time series object. If no such argument is provided as in the third call, a default choice is used.

Figure 1 shows that monitoring of weekly mortality in Denmark requires handling both weekly time series containing small count numbers and series having large counts. For the four age groups in the top row of the figure it will be important to take the count data nature into account, because a Gaussian approximation is expected to work poorly here. As a consequence, we will in our work focus on statistical modeling and aberration detection handling small counts. The methods should, however, be flexible enough to also handle time series with large counts as e.g. in the bottom row of Fig. 1. An additional advantage of being able to handle small counts is that this also allows for further stratification of the

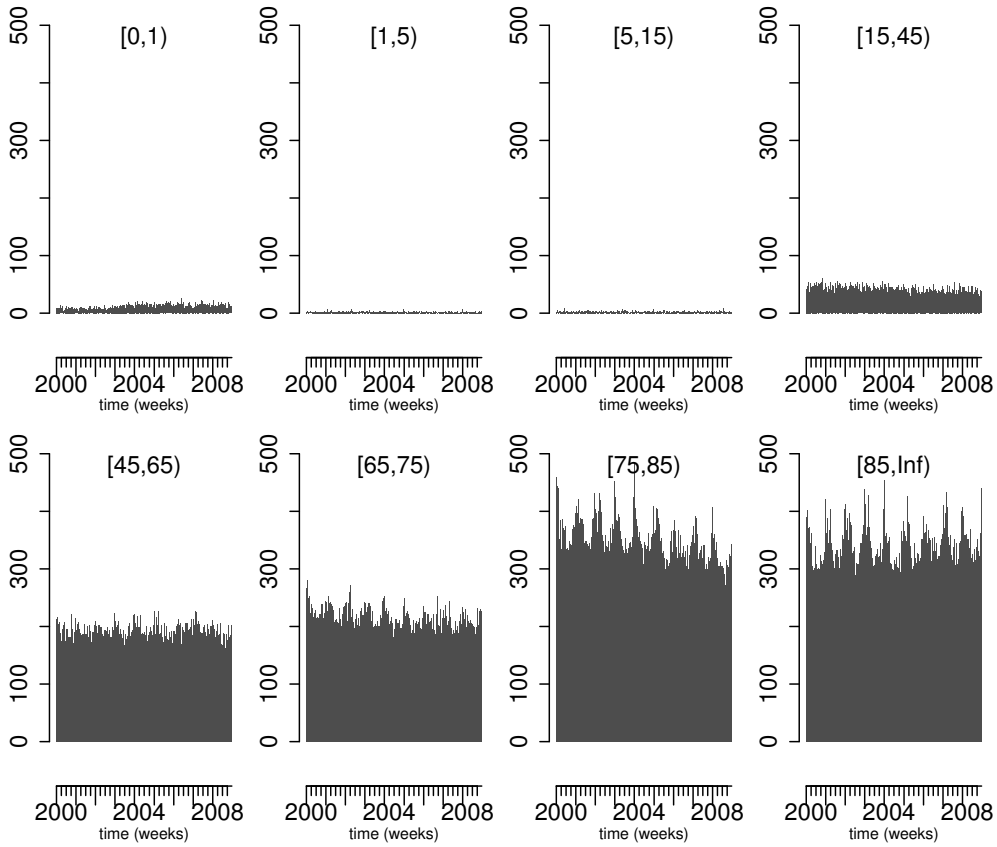


**Figure 1.** Weekly number of all-cause mortality in Denmark in the eight age groups during 2000-2008. Each axis tick denotes a quarter (3-month period) and the larger tick marks denote the 1st quarter of the year (starting with ISO week 01).

data into e.g. geographical regions or gender. Furthermore, the time series can contain temporal trends, e.g. the downward trend for the 65-74 group or the mortality increase in the $\geq 85$ group due to increasing longevity. Similar examples are the seasonal patterns for the 75-84 and $\geq$85 age groups, where an increased mortality during winter and spring is observed. In order to accommodate such non-stationarity we want to investigate modeling and aberration detection approaches taking such trend and seasonality into account.

## 3. Statistical Framework for Aberration Detection

Denote by $\{y_t,\ t = 1, 2, \ldots\}$ the univariate time series to monitor. In this chapter, $y_t$ will always be a discrete univariate random variable, but continuous and multivariate versions are just as conceivable. The aim of aberration detection is to on-line detect an important change in the process occurring at an unknown time $\tau$. This could for example be a change in the process parameters resulting in a change in level or variation of the process. Using terminology from statistical process control, the process can thus be in one of two states

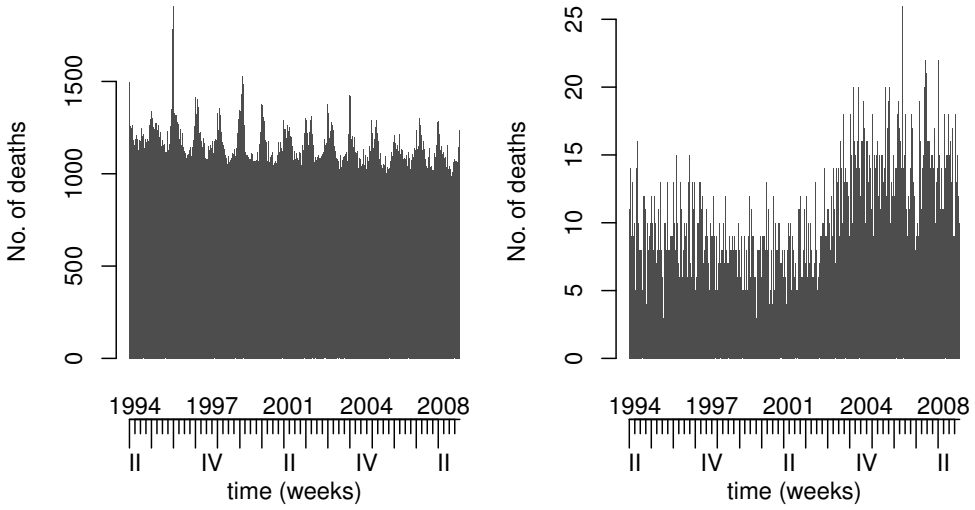


**Figure 2.** (left) Weekly number of all-cause deaths in Denmark 1994-2008 aggregated over all age groups and (right) the weekly data of the <1 age group. The increase of mortality in this agegroup is due to a change in the gestational age defining a stillbirth, which was lowered from 28 weeks to 22 weeks in 2004.

at each time point $t$: *in-control*, i.e. $s < \tau$, or *out-of-control*, i.e. $s \geq \tau$. The binary 0/1 indicator $x(t)$ will denote the true but unknown state of the process at time $t$, assuming that $x(t) = 1$ means out-of-control.

At time $s \geq 1$, where a decision about the state of $x(s)$ is to be made, the available process information is $\boldsymbol{y}_s = \{y_t \,;\, t \leq s\}$. A detection method is now a rule, which predicts the unknown state of $x(s)$ based on $\boldsymbol{y}_s$. This is done by computing a summary $r(\boldsymbol{y}_s)$ based on $\boldsymbol{y}_s$, which is then compared to a threshold value $g$ and consequently

$$\hat{x}(s) = I(r(\boldsymbol{y}_s) > g),$$

where $I(\cdot)$ is an indicator function, i.e. the function returns 1 if $r(\boldsymbol{y}_s) > g$ and zero otherwise. The time of the first out-of-control alarm is then a random variable

$$T_A = \min\{s \geq 1 : r(\boldsymbol{y}_s) > g\}. \tag{1}$$

After the change to the out-of-control state at time $\tau$, the decision rule should as quickly as possible sound an alarm. However, it might take a number of observations after $\tau$ before enough evidence has been collected to do so. Two important target variables for evaluating the performance of a detection method are the *in-control run-length* $T_A|\tau = \infty$, i.e. the number of epochs before the first wrong alarm, and the *out-of-control run-length* $T_A|\tau = 1$, i.e. the number of epochs to detect an already occurred change. Various summaries such as expectation or median can be computed of these run-length variables. Specifically, the expectation of the in-control run-length $\mathrm{E}(T_A|\tau = \infty)$ – known as the *average in-control run-length* or $P(T_A \leq t_a|\tau = \infty)$ – the probability to get a false alarm within the first $t_a$ epochs of the monitoring – are often used as a criterion when evaluating the performance

of a detection method. A more thorough discussion on such criteria can be found e.g. in Frisén (1992).

## 4. The Farrington Algorithm

The aim of the Farrington et al. (1996) algorithm was to develop a robust and fast method applicable for the routine monitoring of weekly reports on infections for many different pathogens at the former Communicable Disease Surveillance Centre (now Health Protection Agency) in the UK. For the current time point $t_0 = (t_0^{\text{week}}, t_0^{\text{year}})$, i.e. week $t_0^{\text{week}}$ in year $t_0^{\text{year}}$, this is done by formulating a statistical algorithm for predicting the observed number of counts $y_{t_0}$. This prediction is based on a subset of the historic data: Centered around the current week $t_0^{\text{week}}$, e.g. week 23, one includes $w$ values to the left and right of that week together with the week itself, e.g. week 21-25 if $w = 2$. This is done for each of the years $t_0^{\text{year}-1}, \ldots, t_0^{\text{year}-b}$. Thus, a total of $b \cdot (2w+1)$ reference values are extracted. Now an overdispersed Poisson generalized linear model (GLM) with log-link is fitted to the reference values. The GLM has the following mean structure:

$$\mathrm{E}(y_t) = \mu_t, \quad \text{where } \log(\mu_t) = \alpha + \beta t, \tag{2}$$

and $\mathrm{Var}(y_t) = \phi\mu_t$ with $\alpha$, $\beta$ and $\phi > 0$ being coefficients to estimate. See e.g. Fahrmeir and Tutz (2001) for further information about GLMs. One can show that an approximate $(1-\alpha)\cdot 100\%$ prediction interval for $y_{t_0}$ based on this GLM has upper limit

$$u_{t_0} = \hat{\mu}_{t_0} + z_{1-\alpha/2} \cdot \sqrt{\mathrm{Var}(y_{t_0} - \hat{\mu}_{t_0})} = \hat{\mu}_{t_0} \cdot \left(1 + z_{1-\alpha/2} \cdot \sqrt{\frac{\hat{\phi}\hat{\mu}_{t_0} + \mathrm{Var}(\hat{\mu}_{t_0})}{\hat{\mu}_{t_0}^2}}\right),$$

where $z_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of the standard normal distribution while $\hat{\mu}_{t_0}, \hat{\phi}$ and $\mathrm{Var}(\hat{\mu}_{t_0})$ can be obtained from the GLM output. If the observed value $y_{t_0}$ is greater than $u_{t_0}$ then the time point $t_0$ is flagged as an outbreak, i.e. in the notation of Section 3:

$$\hat{x}(t_0) = I\left(\frac{y_{t_0}}{u_{t_0}} > 1\right). \tag{3}$$

The Farrington algorithm contains a number of additional refinements for improving the prediction of $y_{t_0}$, for example by correcting for past outbreaks among the reference values, by testing the need of the trend component in (2) and by a skewness correction of the predictive distribution for low count series. In order to keep the current presentation compact, we refer to Farrington et al. (1996) for further details on these refinements. In `surveillance`, the function `farrington` is used to run the algorithm:

```
R> phase2 <- which(epoch(momo) >= "2007-10-01")
R> s.far <- farrington(momo[, "[0,1)"], control = list(range = phase2,
+     alpha = 0.01, b = 5, w = 4, powertrans = "none"))
```

We start the monitoring in week 40 of 2007 (i.e. 1st October 2007) and let `phase2` denote the index of all ISO weeks to monitor. The call to function `farrington` then performs aberration detection for these weeks in the <1 age group. Note that all aberration detection algorithms in `surveillance` follow the same structure: The first argument denotes an object

of class `sts` containing the data, the second argument contains a list of algorithm specific control options and a vector `range` with the time points to monitor. Specifically, the above code uses $\alpha = 0.01$ to form the upper limit of the predictive distribution and $b = 5$ and $w = 4$ to generate the reference values. Figure 3 shows the results of the monitoring. In order to obtain the above described procedure without any additional transformation, the argument `powertrans="none"` is used. Other options are: `"2/3"`, which provides a skewness correction, which is preferable in low count scenario. Similarly, `"1/2"` provides the variance stabilizing square-root transformation.

```
R> plot(s.far, ylab = "No. of deaths", xlab = "time (weeks)", main = "")
```

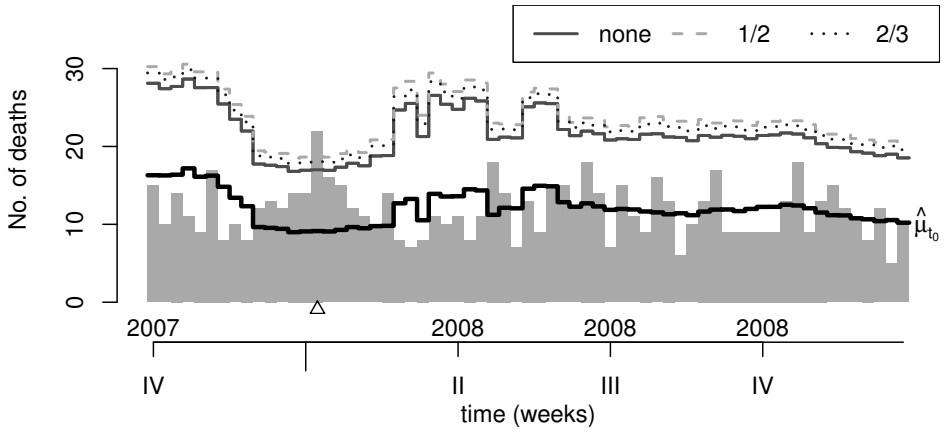


**Figure 3.** Aberration detection for the <1 age group using the Farrington et al. (1996) method. The upper three lines show the upper prediction limit $u_{t_0}$ as calculated using each of the three possible power transformations. The lower solid line denotes the expected model predicted number of cases for each time point $t_0$. Triangles indicate an alarm.

The figure is interpreted as follows: Starting in ISO week 40 of 2007, we use only values from the past to construct a prediction interval for the observed number of counts for week 40. When comparing the actually observed number 15 with the upper limit $u_{t_0} = 28.1$, we have no reason to believe in an excess number of deaths and hence no alarm is generated. The upper limit would have been 30.3 or 29.4 cases for the two other transformations. The same procedure is now repeated for ISO week 41, etc. In week 02 of 2008 the observed number of counts exceeds the threshold of the "none" line for the first time and hence an alarm is generated for that week. No further alarms are generated during the 65 weeks of surveillance. Once an alarm is sounded, the alarm must be verified and the public health significance investigated. In this instance, investigation of available epidemic intelligence did not reveal any specific explanation for the mortality peak that would indicate a significant public health event.

Note also the prospective behavior of the detection: At each time point we are only allowed to look back in time, never ahead in time. Thus detection mimics the arrival of new

data each week, which would be the case in practical applications. Choosing a specific value for $\alpha$ is particularly depending on the application and mode of operation. A value of e.g. $\alpha = 0.01$ means that for a particular week the probability of observing a value $y_{t_0} > u_{t_0}$ by pure chance under the estimated model is $\alpha/2 = 0.5\%$. If these probabilities are assumed independent for the individual weeks, the probability of observing a false alarm during the 65 epochs of the monitoring is thus $1 - (1 - 0.005)^{65} = 0.28$. In section 5.1 we will study in further detail the actual run-length distribution of the algorithm.

A call to an aberration detection algorithm fills the `alarm` slot of the `sts` object. This is an $n' \times m$ matrix of Booleans stating for each time point (aka. epoch) and series whether the time point was classified as aberration. Here, $n'$ corresponds to the number of elements in the `range` argument of the call. Furthermore, the `upperbound` slot contains an $n' \times m$ matrix of values corresponding to the minimum number of cases each week that would have resulted in an alarm. Finally, the slot `control` contains the list of control arguments which was used to invoke the aberration detection algorithm.

For the EuroMOMO project, an important aspect besides the detection of aberrations is the quantification of excess mortality. A first measure of this excess could be based on the predictive distribution. For example, Fig. 3 shows the predicted expected number $\hat{\mu}_t$ of cases in-control, allowing for a definition of excess as e.g. $y_t - \hat{\mu}_t$. By computing confidence intervals for $\hat{\mu}_t$, one would also be able to assess the uncertainty of such an excess. As a further tool in this direction, Fig. 4 shows the quantiles of the predictive distribution. The Farrington procedure sounds an alarm once the $1 - \alpha/2$ quantile is exceeded.

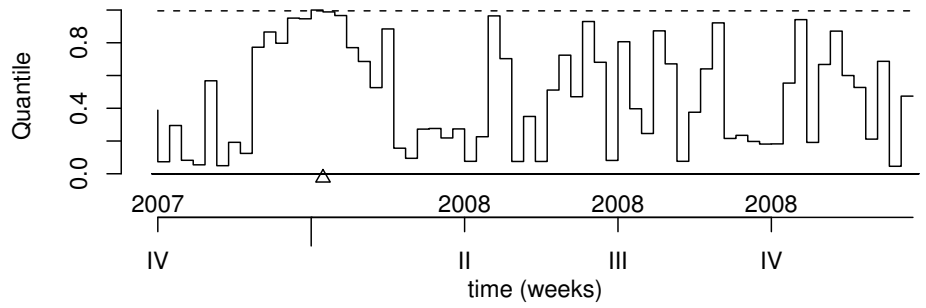


**Figure 4.** Quantiles of the predictive distribution. The dashed line indicates the $1 - \alpha/2 = 0.995$ quantile used for the surveillance in Fig. 3. Triangles indicate the alarms.

One way to simultaneously monitor all eight age groups is to monitor each time series separately using e.g. the Farrington procedure. This is done by the following code:

```
R> s.far.all <- farrington(momo, control = list(range = phase2, alpha = 0.01,
+     b = 5, w = 4))
```

A plot of the alarms for each time series provides a graphical overview as shown in Fig. 5. Monitoring each series independently as done above ignores possible correlations of the time series. Furthermore, if one wanted to keep the number of false alarms at the same level as for the surveillance of a single series, one could, however, have used an $\alpha$ being 1/8 of what was used for the single time series case previously.

```
R> plot(s.far.all, type = alarm ~ time, xlab = "time (weeks)")
```

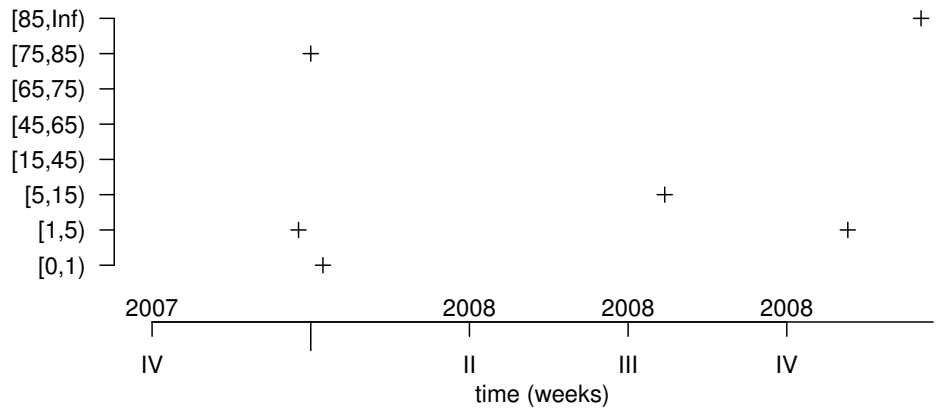


**Figure 5.** Overview of aberration detection for all eight age group time series using the Farrington algorithm with $\alpha = 0.01$.

In Fig. 5 we observe no specific patterns of the alarms across age strata, except for single week alarms around the turn of the year 2007/2008 in age groups <1, 1-4 and 75-84. Note that the 2007/2008 season in Denmark did not exhibit any heavy influenza activity. Furthermore, the current surveillance does correct for any population demographics using the linear trend in (2). It might, however, be worth investigating an additional adjustment for population size in the eight age strata as these are expected to change over the years.

As a further remark, in the notation of Section 3 the Farrington algorithm does not utilize all available information at decision time $s = t_0$, i.e. $r_{\text{Farr}}(\boldsymbol{y}_s) = r_{\text{Farr}}(\boldsymbol{y}'_s)$ with $\boldsymbol{y}'_s \subset \boldsymbol{y}_s$. The effects of seasonality are handled robustly by using only 'similar' weeks as reference values and hence no explicit seasonal model is needed. Such an approach is, however, suboptimal, if it is possible to adequately model the seasonal behavior as e.g. done in Section 5.

Even though more than a single $y_s$ is used to compute $r(\boldsymbol{y}_s)$ in the Farrington algorithm, the decision in (3) occurs by only comparing the current observation with the upper limit of the predictive distribution. Hence, no accumulation of evidence against the in-control situation occurs. In the next section, we reconsider this task from a statistical process control viewpoint and describe an approach taking accumulation into account.

## 5. Negative Binomial CUSUM

Reconsidering (1) more from the viewpoint of statistical process control, the simplest class of detectors is the *Shewhart detector*, which for $r(\boldsymbol{y}_s)$ only utilizes information about the last time point, e.g. by comparing the single $y_s$ value to a fixed threshold value. In a parametric detection setup one assumes a known probability mass function (PMF) $f(\cdot; \boldsymbol{\theta})$ for $\boldsymbol{y}_s$, which is parametrized by a parameter vector $\boldsymbol{\theta}$. If the parameter vector $\boldsymbol{\theta}$ is assumed to be known in the in-control and out-of-control state, an optimal change-point detection can be achieved based on the partial likelihood ratio (Frisén, 2003). Let $L(s, t)$ with $s \geq t$ be the partial likelihood ratio between the out-of-control and in-control models at time $s$ given that $\tau = t$.

Assuming independence between the elements of $\boldsymbol{y}_s$ when conditioning on the parameter $\boldsymbol{\theta}$, one obtains

$$L(s,t) = \frac{f(\boldsymbol{y}_s|\tau = t)}{f(\boldsymbol{y}_s|\tau > s)} = \frac{\prod_{i=1}^{t-1} f(y_i;\boldsymbol{\theta}_0)\prod_{i=t}^{s} f(y_i;\boldsymbol{\theta}_1)}{\prod_{i=1}^{s} f(y_i;\boldsymbol{\theta}_0)} = \prod_{i=t}^{s} \frac{f(y_i;\boldsymbol{\theta}_1)}{f(y_i;\boldsymbol{\theta}_0)}.$$

For the Shewhart detector, optimal detection can be achieved by $r(s) = L(s,s)$. This detector is good at detecting large process shifts quickly, but if the shift is small but sustained, accumulating deviations over time is necessary in order to detect the change. The likelihood ratio based cumulative sum originally proposed by Page (1954) is one method to deal with accumulation and is advantageous for detecting sustained shifts. It uses

$$r(s) = \max\{1 \leq t \leq s : L(s,t)\}.$$

When the $y_t$ are independent and identically distributed discrete random variables, such count data CUSUM detectors are well investigated, see e.g. Hawkins and Olwell (1998). However, biosurveillance data often exhibit seasonal variations and time trends which violate the assumption of an identical distribution. As in Höhle et al. (2009), let

$$r(s) = \max_{1\leq t\leq s}\left[\sum_{i=t}^{s} \log\left\{\frac{f(y_i;\boldsymbol{\theta}_1)}{f(y_i;\boldsymbol{\theta}_0)}\right\}\right], \tag{4}$$

where we have used the loglikelihood ratio (LLR) instead of the likelihood ratio. Let $\boldsymbol{\theta}_0$ denote the in-control and $\boldsymbol{\theta}_1$ the out-of-control parameters. If $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$ are known, (4) can be written in recursive form as follows:

$$r_0 = 0 \quad \text{and} \quad r_s = \max\left(0, r_{s-1} + \log\left\{\frac{f(y_s;\boldsymbol{\theta}_1)}{f(y_s;\boldsymbol{\theta}_0)}\right\}\right), \quad \text{for } s \geq 1. \tag{5}$$

One sees that for time points with LLR>0, i.e. evidence against in-control, the LLR contributions are added up. On the other hand, no credit in the direction of the in-control is given because $r_s$ cannot get below zero.

In practical applications, the in-control and out-of-control parameters are, however, hardly ever known beforehand. A typical procedure in this case is to use historical *phase 1 data* for the estimation of $\boldsymbol{\theta}_0$ with the assumption that these data originate from the in-control state. This estimate is then used as plug-in value in the above LLR. Simultaneously, the out-of-control parameter $\boldsymbol{\theta}_1$ is specified as a known function of $\boldsymbol{\theta}_0$, e.g. as a known multiplicative increase in the mean. Developing appropriate count data time series models together with statistical inference for the estimation of $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$ in a statistical process control framework is thus an important aspect of performing biosurveillance.

As we suspect the number of persons in the eight age groups to shift towards older age during the years, we want to take the population size of the eight age-strata into account in our monitoring. We do so by using data from Statistics Denmark (2009) on the number of individuals on 1st Jan 1994-2008 in each of the eight age groups.

```
R> population(momo) <- as.matrix(read.csv("population-dk.csv", check.names = FALSE))
```

We will in the following use a generalized log-link negative binomial model for the in-control situation of a specific age group, i.e. $y_t \sim \text{NegBin}(\mu_{0,t}, \alpha)$ with

$$\log(\mu_{0,t}) = \beta_0 + \beta_1 \cdot t + c(t) + \beta_2 \cdot \text{pop}_t, \tag{6}$$

where $c(t)$ is a cyclic function with period 52 or 53 depending on the number of ISO weeks in the year of $t$, e.g. $c(0) = c(52)$ for years with 52 ISO weeks. Such behavior can e.g. be obtained by sinusoidals as in Serfling (1963) or using cyclic splines (Wood, 2006). Furthermore, $\text{pop}_t$ denotes the population size in the respective age group at time $t$. In the above negative binomial model $\mathrm{E}(y_t) = \mu_{0,t}$ and $\mathrm{Var}(y_t) = \mu_{0,t} + \alpha \cdot \mu_{0,t}^2$, i.e. $\alpha$ is a dispersion parameter, which we will assume to be constant over time. Thus with $\alpha > 0$ we are able to handle possible overdispersion of the count data time series. For $\alpha \to 0$ the negative binomial distribution tends to the Poisson distribution.

The out-of-control model for the mean is now assumed to be $\mu_{1,t} = \kappa \cdot \mu_{0,t}$ which on the log-link scale corresponds to a level shift in the intercept from $\beta_0$ to $\beta_0 + \log(\kappa)$. The following R code estimates such a negative binomial GLM from the phase 1 data of the 75-84 age group using the `glm.nb` function (Venables and Ripley, 2002).

```
R> phase1 <- which(year(momo) == 2002 & epochInYear(momo) == 40):(phase2[1] -
+     1)
R> momo.df <- as.data.frame(momo)
R> m <- glm.nb(`observed.[75,85)` ~ 1 + epoch + sin(2 * pi * epochInPeriod) +
+     cos(2 * pi * epochInPeriod) + `population.[75,85)`, data = momo.df[phase1,
+     ])
R> mu0 <- predict(m, newdata = momo.df[phase2, ], type = "response")
```

Here, `phase1` contains the index of all time epochs in the phase 1 sample used to estimate the in-control parameters. A five-year period has been used above. Then the function `as.data.frame` is applied to convert the `sts` object to the necessary `data.frame` used by `glm.nb`. For simplicity, a single harmonic is used for $c(t)$ consisting of one sine and one cosine term. The parameter estimates for the other terms are $\hat{\beta}_0 = 10.49$, $\hat{\beta}_1 = -9.54 \cdot 10^{-5}$, $\hat{\beta}_2 = -1.24 \cdot 10^{-5}$ and $\hat{\alpha} = 1.97 \cdot 10^{-3}$. In practical application one should perform a model selection process to decide on covariates and an appropriate number of harmonics to include. For example, such a selection for the above model would reveal $\text{pop}_t$ as being non-significant, whereas a total of three superimposed harmonics could be justified. For illustration we, however, proceed with the above model and use `predict` to obtain the expected value $\mu_{0,t}$ during phase 2. Figure 6 illustrates the $\mu_{0,t}$ predictions based on this GLM model.

If, for example, one wanted to optimally detect a 20% increase in the mean, one would have $\kappa = 1.2$. Again, the choice of $\kappa$ depends very much on the specific application and mode of operation. Together with the threshold $g$ the value of $\kappa$ determines the distribution of the run-length as further investigated in Section 5.1. The resulting $\mu_{1,t}$ is shown in Fig. 6. Also shown is the number needed before alarm (NNBA) at each time $s$. This number is obtained by reversing (5) with known threshold $g$, i.e. given $r_{s-1}$ find the minimum $y_s$ such that $r_s > g$.

```
R> kappa <- 1.2
R> s.nb <- glrnb(momo[, "[75,85)"], control = list(range = phase2,
+     alpha = 1/m$theta, mu0 = mu0, c.ARL = 4.75, theta = log(kappa),
+     ret = "cases"))
```

The above code extracts the dispersion parameter $\alpha$ from the `glm.nb` fit – note the slightly different parametrization of the dispersion parameter here. For the threshold $g$ (in `glrnb` denoted `c.ARL`) we use the value of 4.75. This threshold value determines the distribution of the run-length $T_A$ as investigated in detail in Section 5.1 – specifically we show that

$g \approx 4.75$ results in $P(T_A \leq 65|\tau = \infty) \approx 0.1$. The results from this call are illustrated in Fig. 6. For week 02 in 2008 an alarm is generated. Notice that the number of cases in the previous week is not enough to sound an alarm itself, but helps to lower the NNBA in the following week, where it is just about exceeded. No further alarms are generated. The alarm is an example of excess mortality peaks in the elderly that occur regularly during winter around the change of the year and at the time of influenza epidemics.

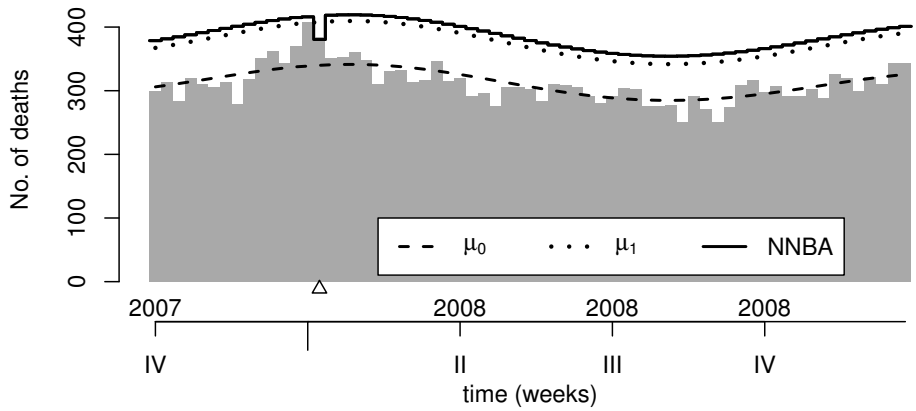


**Figure 6.** Aberration detection for the 75-84 age group series using a time varying negative binomial CUSUM. Shown are the time varying in-control and out-of-control means and the number needed before an alarm (NNBA). Triangles indicate alarms.

### *5.1. Run-length properties*

As mentioned previously, the behavior of the CUSUM depends very much on the choice of the threshold $g$. In order to guide the choice of $g$, we will look at the run-length distribution of $T_A|\tau = \infty$ under the fitted negative binomial model. Prediction of $\mu_{0,t}$ requires knowledge of all involved covariates during the monitoring period, e.g. in model (6) this would be the population size. For the monitored period of 65 weeks (2007-W40 – 2008-W52), these values are available, but if monitoring exceeded this period, we would have needed to predict covariate values as well before being able to compute $\mu_{0,t}$. Hence, it is practically more feasible to look at $P(T_A \leq t_A|\tau = \infty)$ for a small $t_A$ than to estimate e.g. $\mathrm{E}(T_A|\tau = \infty)$ as here many more time points might be needed if the expectation is large. Furthermore, the distribution of $T_A$ is also often skew, which makes the expectation a bad summary of the central tendency.

Specifically, we want to choose $g$ such that $P(T_A \leq 65|\tau = \infty)$ is below some acceptable value, e.g. 10%. In other words, the probability of a false alarm within the 65 weeks of our $\mu_{0,t}$ vs. $\mu_{1,t}$ monitoring should be below 10%. To compute the probability under the selected model, two approaches exist: direct Monte Carlo estimation or a Markov chain

approximation.

In the first approach, we use Monte Carlo estimation of $P(T_A \leq 65|\tau = \infty)$. For each realization $j$, a time series of length 65 is simulated from the estimated negative binomial model with mean $\mu_{0,t}, t = 1, \ldots, 65$, and dispersion parameter $\alpha$. Then the negative binomial CUSUM is applied to this time series and one checks if $T_A^j \leq 65$. The probability of interest using $k$ such realizations can then be estimated as $\sum_{j=1}^{k} I(T_A^j \leq 65)/k$, where $I(\cdot)$ is the indicator function. Code-wise this can be done for $k = 1000$ for a grid of $g$'s as follows.

```
R> simone.TAleq65 <- function(sts, g) {
+     observed(sts)[phase2, ] <- rnbinom(length(mu0), mu = mu0, size = m$theta)
+     one <- glrnb(sts, control = modifyList(control(s.nb), list(c.ARL = g)))
+     return(any(alarms(one)))
+ }
R> g.grid <- seq(1, 8, by = 0.5)
R> pMC <- sapply(g.grid, function(g) {
+     mean(replicate(1000, simone.TAleq65(momo[, "[75,85)"], g)))
+ })
```

Figure 7 shows the result. We note that $g \approx 4.75$ ensures that the false-alarm probability within the monitoring period drops below the desired level of 10%. If one is interested in $P(T_A \leq 65|\tau = 1)$ instead, $\mu_{1,t}$ has to be used as argument `mu` in `rnbinom`.

A different option to compute the above false-alarm probability for a likelihood ratio based CUSUM is to use a Markov chain approximation to determine the PMF of the run-length variable. This approach implemented in `surveillance` is a generalization of the work in Bissell (1984) to time varying count data CUSUMs.

```
R> dY <- function(y, mu, log = FALSE, alpha, ...) {
+     dnbinom(y, mu = mu, size = 1/alpha, log = log)
+ }
R> pMarkovChain <- sapply(g.grid, function(g) {
+     TA <- LRCUSUM.runlength(mu = t(mu0), mu0 = t(mu0), mu1 = kappa *
+         t(mu0), h = g, dfun = dY, n = rep(600, length(mu0)), alpha = 1/m$theta)
+     return(tail(TA$cdf, n = 1))
+ })
```

Here, `dY` is a function specifying the one-parameter PMF used in the likelihood ratio detector, in our case this is the negative binomial PMF $f(y_t; \mu_t, \alpha)$. The above invocation of the function `LRCUSUM.runlength` derives the distribution of $T_A$ when the value of $\mu_t$ is equal to $\mu_{0,t}$ (i.e. in-control) for given specifications of in-control mean, out-of-control mean and dispersion parameter. The function computes the loglikelihood ratio between all possible realizations of $y_t$. However, to make computations feasible, an upper limit `n` is used at each time point, after which for $y_t > n$ the probability of $y_t$ to occur under $\mu_t$ is negligible. Figure 7 shows the result and the close agreement with the Monte Carlo estimation. The Markov chain approximation is considerably faster though.

Returning to the monitoring of the <1 age group from Section 4, we would like to compare the Farrington algorithm with the negative binomial CUSUM. To do so, we use the in-control model $\text{NegBin}(\mu_{0,t}, \alpha)$ for the CUSUM, with $\mu_{0,t}$ as in Fig. 3 and $\alpha$ estimated by a similar GLM as in Section 5. The out-of-control mean is again given as $\mu_{1,t} = 1.2 \cdot \mu_{0,t}$. The threshold $g$ should be chosen such that the two algorithms are as comparable as possible

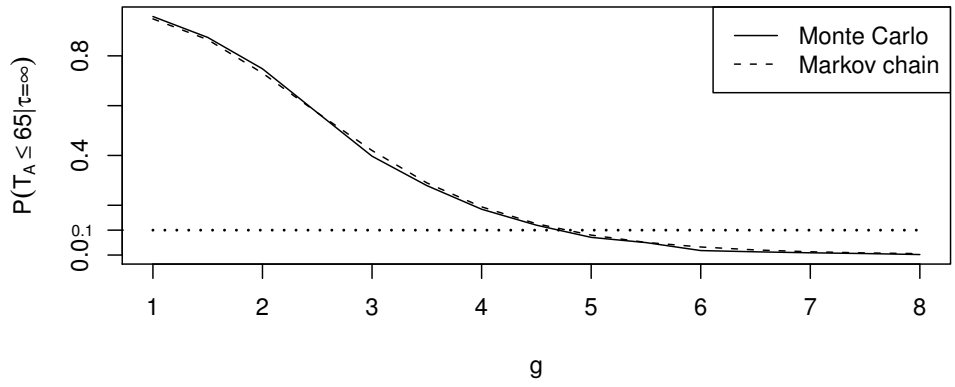


**Figure 7.** $P(T_A \leq 65|\tau = \infty)$ as a function of the threshold $g$ computed using both Monte Carlo simulation and a Markov chain approximation. A dotted line shows the desired value 0.10.

with respect to e.g. $P(T_A \leq 65|\tau = \infty)$. A Monte Carlo estimation as just described is performed to determine this probability for the Farrington algorithm. The model used for this simulation is the above in-control negative binomial model.

Based on 1000 realizations of $I(T_A \leq 65|\tau = \infty)$, the probability is estimated to be 0.57, which is surprisingly high compared to the rough estimate of 0.28 in Section 4. However, the two numbers are not completely comparable as the simulation uses a negative binomial model and observations are not independent. If the above Monte Carlo estimated false alarm probability of the Farrington algorithm should be near 10%, we would have to choose a much smaller $\alpha$. Instead we use the Markov chain approximation to determine that a threshold of $g \approx 2.2$ gives a similar probability for the negative binomial CUSUM. Figure 8 contains the result of the CUSUM monitoring with this threshold.

The CUSUM behaves slightly different than the Farrington algorithm in Fig. 3. In the last weeks of 2007, an increased number of cases above the baseline is accumulated leading to a steady decrease of NNBA. In week 01, the threshold is nearly reached, but as for the Farrington procedure, an alarm is first generated for week 02 in 2008. However, the sustained excess above baseline leads to a further alarm in week 08, which was not detected by the Farrington algorithm, as here, the excess alone in that week is not enough to get beyond the threshold.

## 6. Discussion

In this chapter we have given an introduction to the capabilities of the open-source R package `surveillance` for epidemiological biosurveillance. Further advantages of choosing R to conduct such analyses exist: R produces high-quality graphics in a variety of formats, including TIFF, PNG, EPS and PDF which combined with Sweave or odfWeave (Leisch, 2002; Kuhn and Weaston, 2009) allows for automatic report generation using LaTeX/OpenOffice in literate programming fashion. Also HTML pages containing text, graphics and tables of the results can automatically be generated from R using e.g. the package `R2HTML` (Lecoutre, 2003) or `hwriter` (Pau, 2009). Altogether, using the command

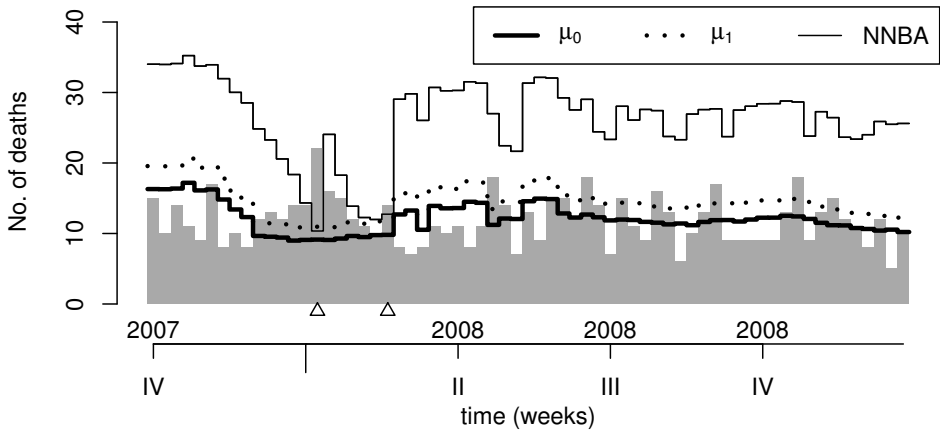


**Figure 8.** Negative binomial CUSUM for the <1 group. The interpretation of the lines shown is as in Fig. 6.

```
R> demo(biosurvbook)
```

the analyses of this chapter can be reproduced after the package has been loaded.

We introduced the time varying negative binomial CUSUM as an alternative to the Farrington aberration detection method, as it is better embedded within the framework of statistical process control. A shortcoming of the suggested GLM modeling to determine in-control and out-of-control values is that any uncertainty of the estimation was ignored when plugging in the estimators for $\mu_{0,t}$ and $\mu_{1,t}$ into the CUSUM. Furthermore, no auto-correlation between observations was taken into account – neither in the GLM model nor in the likelihood ratio based CUSUM. However, if trend and seasonality are adequately modeled, little auto-correlation is expected to remain as e.g. shown in the simulation study by Farrington et al. (1996). If auto-correlation is a concern, different modeling strategies can be applied; e.g. generalized estimating equations (used for mortality modeling in e.g. Fouillet et al. (2008)), integer auto-regressive models (Freeland and McCabe, 2004; Held et al., 2005; Weiß, 2007) or pairwise likelihood models (Varin and Vidoni, 2006). An auto-regressive approach is e.g. implemented in the function `glrnb` by using the control argument `change="epi"`, see Höhle and Paul (2008) for details on the methodology. The same reference also discusses how to estimate the out-of-control state at each time point using generalized likelihood ratio CUSUMs instead of a fixed prior specification. An alternative to the independence assuming likelihood ratio based CUSUM is the Shiryaev-Roberts detector, which also works for auto-correlated observations, see e.g. Frisén (2003) for details. As an example, the spatio-temporal cluster detection of Assunçáo and Correa (2009) – implemented as function `sr` in `surveillance` – uses this detector. Further package developments are the extension to categorical time series, e.g. the monitoring of binomial and multinomial data.

With respect to the Danish mortality monitoring, the presented analyses illustrated the potential of using `surveillance` and R for this task since they provide methods for the visualization, modeling and aberration detection. A big advantage of the regression based models in the CUSUM detection is their flexibility for extending them with additional covariates as illustrated by population size. Such covariates could e.g. be the number of influenza like illness cases or temperature. A limitation of the current methods is that mortality reporting is governed by a delay between the day of death and the reporting to health authorities. Quantification and handling of such reporting delay is thus a precondition for valuable prospective monitoring. Approaches exist for dealing with such reporting delay, see e.g. Heisterkamp et al. (2006), but these are currently not available for routine use in `surveillance` and have also not methodologically been adapted to the CUSUM context. Finally, the open-source and copyleft approach of the R system and `surveillance` is well suited for the EuroMOMO project aim of obtaining a mortality monitoring system operating in many different countries.

### *6.1. Acknowledgements*

We thank Christiane Dargatz for her proofreading of the manuscript. The EuroMOMO project has received funding from the European Union in the framework of the Public Health Programme. However, the responsibility for the content lies solely with the authors and the European Union is not responsible for any use that may be made of the information.

## References

Anonymous (2004, 12 Jan). Data elements and interchange formats – information interchange – representation of dates and times. Technical Report ISO 8601:2004(E), The International Organization for Standardization.

Anonymous (2009). European monitoring of excess mortality for public health action. Project homepage: `http://www.euromomo.eu/`, last access: 9 June 2009.

Aragon, T. (2008). *epitools: Epidemiology Tools*. R package version 0.5-2.

Assunçáo, R. and T. Correa (2009). Surveillance to detect emerging space-time clusters. *Computational Statistics and Data Analysis 53*(8), 2817–2830.

Bissell, A. F. (1984). The performance of control charts and cusums under linear trend. *Applied Statistics 33*(2), 145–151.

Bravata, D. M., K. M. McDonald, W. M. Smith, C. Rydzak, H. Szeto, D. L. Buckeridge, C. Haberland, and D. K. Owens (2004). Systematic review: Surveillance systems for early detection of bioterrorism-related diseases. *Annals of Internal Medicine 140*(11), 910–922.

Buckeridge, D., H. Burkom, M. Campbell, W. Hogan, and A. Moore (2005). Algorithms for rapid outbreak detection: a research synthesis. *Journal of Biomedical Informatics 38*(2), 99–113.

Conti, S., P. Meli, G. Minelli, R. Solimini, V. Toccaceli, M. Vichi, C. Beltrano, and L. Perini (2005). Epidemiologic study of mortality during the summer 2003 heat wave in Italy. *Environmental Research 98*(3), 390–399.

Dalgaard, P. (2008). *Introductory Statistics with R* (2nd ed.). Springer.

Fahrmeir, L. and G. Tutz (2001). *Multivariate Statistical Modelling Based on Generalized Linear Models* (2nd ed.). Springer.

Farrington, C., N. Andrews, A. Beale, and M. Catchpole (1996). A statistical algorithm for the early detection of outbreaks of infectious disease. *Journal of the Royal Statistical Society, Series A 159*, 547–563.

Fouillet, A., V. W. G. Rey, K. Laaidi, P. Empereur-Bissonnet, A. Le Tertre, P. Frayssinet, P. Bessemoulin, F. Laurent, P. De Crouy-Chanel, E. Jougla, and D. Hémon (2008). Has the impact of heat waves on mortality changed in France since the European heat wave of summer 2003? A study of the 2006 heat wave. *International Journal of Epidemiology 37*, 309–317.

Freeland, R. K. and B. P. M. McCabe (2004). Analysis of low count time series data by Poisson autoregression. *Journal of time series analysis 25*(5), 701–722.

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*(2), 403–434.

Gergonne, B., R. Pebody, N. Andrews, W. Bird, M. Gibbs, and C. Griffiths (2005). *Mortality early warning system based on weekly death registrations in England*, Chapter Oral presentation 154, pp. 107. Annual Conference, Warwick, UK, 12-14 September 2005, Abstracts: oral presentations and posters. London: Health Protection Agency.

Hawkins, D. M. and D. H. Olwell (1998). *Cumulative Sum Charts and Charting for Quality Improvement*. Statistics for Engineering and Physical Science. Springer.

Heisterkamp, S. H., A. L. M. Dekkers, and J. C. M. Heijne (2006). Automated detection of infectious disease outbreaks: hierarchical time series models. *Statistics in Medcine 25*, 4179–4196.

Held, L., M. Hofmann, M. Höhle, and V. Schmid (2006). A two component model for counts of infectious diseases. *Biostatistics 7*, 422–437.

Held, L., M. Höhle, and M. Hofmann (2005). A statistical framework for the analysis of multivariate infectious disease surveillance data. *Statistical Modelling 5*, 187–199.

Höhle, M. (2007). surveillance: An R package for the monitoring of infectious diseases. *Computational Statistics 22*(4), 571–582.

Höhle, M. and M. Paul (2008). Count data regression charts for the monitoring of surveillance time series. *Computational Statistics and Data Analysis 52*(9), 4357–4368.

Höhle, M., M. Paul, and L. Held (2009). Statistical approaches to the surveillance of infectious diseases for veterinary public health. *Preventive Veterinary Medicine 91*(1), 2–10.

Josseran, L., J. Nicolau, N. Caillère, P. Astagneau, and G. Brücker (2006, 01 Dec). Syndromic surveillance based on emergency department activity and crude mortality: two examples. *Eurosurveillance 11*(12), pii=668.

Kaiser, R., D. Coulombier, M. Baldari, D. Morgan, and C. Paquet (2006, 02 Feb). What is epidemic intelligence, and how is it being improved in Europe? *Eurosurveillance 11*(5), pii=2892.

Kuhn, M. and S. Weaston (2009). *odfWeave: Sweave processing of Open Document Format (ODF) files*. R package version 0.7.10.

Le Strat, Y. and F. Carrat (1999). Monitoring epidemiologic surveillance data using hidden Markov models. *Statistics in Medicine 18*, 3463–3478.

Lecoutre, E. (2003, December). The R2HTML package. *R News 3*(3), 33–36.

Leisch, F. (2002). *Sweave: Dynamic generation of statistical reports using literate data analysis*, pp. 575–580. Compstat 2002 - Proceedings in Computational Statistics.

Lopez-Quílez, V. G.-R. J. F.-F. A. (2005). Detecting clusters of disease with R. *Journal of Geographical Systems 7*(2), 189–206.

Mazick, A. (2007, 04 Jan). Monitoring excess mortality for public health action: potential for a future European network. *Eurosurveillance 12*(1), pii=3107.

Muenchen, R. A. (2009). *R for SAS and SPSS Users*. Springer.

Nogueira, P. J., J. M. Falcão, M. T. Contreiras, E. Paixão, J. Brandão, and I. Batista (2005, 01 Jul). Mortality in Portugal associated with the heat wave of August 2003: Early estimation of effect, using a rapid method. *Eurosurveillance 10*(7), pii=553.

Page, E. (1954). Control charts for the mean of a normal population. *Journal of the Royal Statistical Society, Series B 16*(1), 131–135.

Pau, G. (2009). *hwriter: HTML Writer - Outputs R objects in HTML format*. R package version 1.1.

Paul, M., L. Held, and A. M. Toschke (2008). Multivariate modelling of infectious disease surveillance data. *Statistics in Medicine 27*, 6250–6267.

R Development Core Team (2009). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org.

Robert Koch Institute (2009). SurvStat@RKI. http://www3.rki.de/SurvStat. Last access: 9 June 2009.

Rogerson, P. and I. Yamada (2004). Approaches to syndromic surveillance when data consist of small regional counts. *Morbidity and Mortality Weekly Report 53*, 79–85.

Rossi, G., L. Lampugnani, and M. Marchi (1999). An approximate CUSUM procedure for surveillance of health events. *Statistics in Medicine 18*, 2111–2122.

Serfling, R. (1963). Methods for current statistical analysis of excess pneumonia-influenza deaths. *Public Health Reports 78*, 494–506.

Simón, F., G. Lopez-Abente, E. Ballester, and F. Martínez (2005, 01 Jul). Mortality in Spain during the heat waves of summer 2003. *Eurosurveillance 10*(7), pii=555.

Sonesson, C. and D. Bock (2003). A review and discussion of prospective statistical surveillance in public health. *Journal of the Royal Statistical Society, Series A 166*(1), 5–12.

Statistics Denmark (2009). Statbank Denmark. Available from http://statbank.dk. Date of query: 6 June 2009.

Stroup, D., G. Williamson, J. Herndon, and J. Karon (1989). Detection of aberrations in the occurrence of notifiable diseases surveillance data. *Statistics in Medicine 8*, 323–329.

Tennant, R., M. A. Mohammed, J. J. Coleman, and U. Martin (2007). Monitoring patients using control charts: a systematic review. *International Journal of Quality in Health Care 19*(4), 187–194.

Varin, C. and P. Vidoni (2006). Pairwise likelihood inference for ordinal categorical time series. *Computational Statistics and Data Analysis 51*, 2365–2373.

Venables, W. N. and B. D. Ripley (2002). *Modern Applied Statistics with S* (4th ed.). Springer.

Venables, W. N., D. M. Smith, and the R Development Core Team (2009). *An Introduction to R* (2.9.0 ed.). The R Project for Statistical Computing. Available from `http://cran.r-project.org/manuals.html`.

Weiß, C. H. (2007). Controlling correlated processes of poisson counts. *Quality and reliability engineering international 23*, 741–754.

Wood, S. (2006). *Generalized Additive Models: An Introduction with R*. Chapman & Hall/CRC.

Woodall, W. (2006). The use of control charts in health-care and public-health surveillance. *Journal of Quality Technology 38*(2), 89–104.

Zeileis, A., F. Leisch, K. Hornik, and C. Kleiber (2002). strucchange: An R package for testing for structural change in linear regression models. *Journal of Statistical Software 7*(2), 1–38.