J. R. Statist. Soc. A (1996)
159, *Part* 3, *pp.* 547–563

A Statistical Algorithm for the Early Detection of Outbreaks of Infectious Disease

By C. P. FARRINGTON† and N. J. ANDREWS and A. D. BEALE and M. A. CATCHPOLE

Public Health Laboratory Service, London, UK *Communicable Disease Surveillance Centre, London, UK*

[Received August 1995. Revised March 1996]

SUMMARY

Outbreaks of infectious diseases must be detected early for effective control measures to be introduced. When dealing with large amounts of data, automated procedures can usefully supplement traditional surveillance methods, provided that the wide variety of patterns and frequencies of infections are taken into account. This paper describes a robust system developed to process weekly reports of infections received at the Communicable Disease Surveillance Centre. A simple regression algorithm is used to calculate suitable thresholds. Organisms exceeding their threshold are then flagged for further investigation.

Keywords: CLUSTERING; DISPERSION; EPIDEMIOLOGY; EXCEEDANCE; GLIM; OUTBREAK; REGRESSION; REWEIGHTING; SEASONALITY; THRESHOLD; TREND

1. INTRODUCTION

Epidemiological surveillance is the routine process of collection, analysis and dissemination of health data for public health purposes (Langmuir, 1963). One of the functions of infectious disease surveillance is to detect outbreaks and to initiate timely interventions. These interventions may include the identification and removal of contaminated food-stuffs, for food-borne infections such as *Salmonella* or *Campylobacter*, vaccination or prophylactic treatment of individuals at risk, e.g. in outbreaks of measles or meningococcal meningitis, or changes in procedures, e.g. water treatment in the case of water-borne infections like *Cryptosporidium*. In England and Wales, the surveillance of infectious diseases is co-ordinated at the national level by the Communicable Disease Surveillance Centre (CDSC). This paper describes a system developed to assist in the detection of outbreaks by automatically scanning the weekly reports of infectious disease received at the CDSC.

Data from surveillance systems, especially those based on voluntary reporting, are often subject to bias and delays. Using such data for the early detection of outbreaks is particularly problematic, since reports must be reviewed and acted on as they accumulate, without the opportunity to correct errors or to adjust for delays in reporting and other artefacts of the reporting system. In this sense, within a notional spectrum of 'statistical correctness', the application of statistical methodology to the detection of outbreaks lies at the opposite extreme to designed experiments and randomized controlled clinical trials. Nevertheless, as this paper demonstrates, a

†*Address for correspondence*: Statistics Unit, Public Health Laboratory Service, 61 Colindale Avenue, London, NW9 5EQ, UK.
E-mail: pfarring@phls.co.uk

 0035–9238/96/159547

statistical approach can usefully supplement other expert procedures, in particular when a large volume of data must be processed rapidly.

Most statistical work for detecting temporal patterns in epidemiology has been motivated by retrospective analyses of case series. For example, Ederer *et al.* (1964) proposed a test statistic based on the maximum number of cases in disjoint time intervals to identify leukaemia clusters. Naus (1965) developed a continuous version, the scan statistic, based on the maximum number of events in a time window of predetermined length as it scans a given time period. This statistic has been widely used in epidemiology (Wallenstein, 1980). Tango (1984) proposed an index of temporal clustering for use with data grouped in discrete time periods. This is defined as $\mathbf{r}^T\mathbf{A}\mathbf{r}$ where $\mathbf{r}$ is the vector of relative frequencies and $\mathbf{A}$ is an arbitrary fixed matrix, a_{ij} representing the closeness of the ith and jth time intervals. Regression techniques have been used retrospectively to identify temporal clustering of health events by using normal (Serfling, 1963) or Poisson (Parker, 1989) errors. Other methods have also been developed for the detection of specific clustering patterns such as seasonality (Roger, 1977).

In the prospective context the aim is to detect clustering of events at one extremity of the data series, typically in the most recent time interval. Thus interest focuses on the distribution of counts at time t, given the history of the process up to t. Most of the techniques mentioned above can also be used prospectively, although methods for detecting non-specific clustering patterns such as the scan statistic or Tango's index may be expected to lack power in this context. However, in contrast with the many references on the retrospective identification of temporal clusters, there appear to have been fewer applications of statistical methods to prospective detection. This problem bears some similarity to quality control in an industrial setting. Thus cumulative sum statistics, originally developed for industrial applications (Page, 1954), have been used to detect the onset of influenza epidemics (Tillett and Spencer, 1982) and changes in the frequency of congenital malformations (Weatherall and Haskey, 1976). Chen *et al.* (1993) proposed a method based on the time intervals between successive events, applicable to very rare events. Stroup *et al.* (1989) described a system for detecting aberrations in reports of notifiable diseases in the USA based on observed over expected ratios, threshold values being derived by parametric methods based on the normal distribution. Time series techniques have also been used for detecting outbreaks (Watier and Richardson, 1991). A different approach using exponential smoothing of the time series identifies the points at which the first derivative of the series departs significantly from 0 (Nobre and Stroup, 1994).

The primary purpose of the application described in this paper is to identify outbreaks sufficiently early to allow time for interventions. The setting and operational requirements of this early warning mechanism, and the constraints on the choices of statistical methodology imposed by the need to process large quantities of data using automated procedures, are discussed in Section 2. The implementation of the system is further complicated by delays in reporting, fluctuations in the historical data series due to seasonal cycles and secular trends, and by past outbreaks, which vary from organism to organism. In addition, the system must be sufficiently robust to handle a wide range of organisms, from uncommon salmonellas with a weekly frequency of less than 1 to organisms such as rotavirus with weekly frequencies of several hundred. The statistical methods applied to cope with these difficulties are

discussed in Section 3. An evaluation of the system using real and simulated data is described in Section 4, and the paper ends with a brief discussion.

2. COMMUNICABLE DISEASE REPORTS AND DETECTION OF OUTBREAKS IN ENGLAND AND WALES

2.1. *Communicable Disease Report Network*

One of the principal mechanisms for national surveillance of infectious disease in England and Wales is the communicable disease report, a laboratory-based passive reporting system. The back-bone of the system is a network of laboratories, including the 49 laboratories of the Public Health Laboratory Service and over 200 National Health Service laboratories, which report to the CDSC in London. Each week, these laboratories send details of bacterial, viral and other organisms identified from specimens submitted for diagnosis by general practitioners and hospital departments. The frequency of reports varies considerably between organisms, as shown in Table 1. Each week, about 200–350 different organism types are reported. Some organisms of particular public health importance, such as *Mycobacterium tuberculosis*, influenza and certain salmonellas, are subject to special scrutiny, and for these special surveillance mechanisms complement the laboratory reporting system. However, the task of reviewing current and past reports for the remaining organisms to detect evolving outbreaks outstrips the capabilities for manual scanning. An automated system for the detection of potential temporal clusters is therefore required to assist in the identification of national outbreaks.

2.2. *What Events should be Detected?*

Many organisms undergo considerable seasonal fluctuations, which may peak at different times of the year, whereas others also display long-term trends (Fig. 1). However, seasonal variability and secular trends are not of primary interest, the aim being to detect increases on top of expected patterns. Furthermore, some unusual fluctuations may not be of epidemiological interest: sporadic cases of a rare organism constitute unusual events in a statistical sense but do not constitute outbreaks. In addition, interest is focused primarily on abnormally high rather than abnormally low counts.

TABLE 1
Distribution of organisms reported to the CDSC in 1994 by frequency of reports

Average weekly frequency	*No. of organisms*
>250	4
101–250	7
21–100	20
11–20	13
6–10	30
1–5	1129
<1	2725

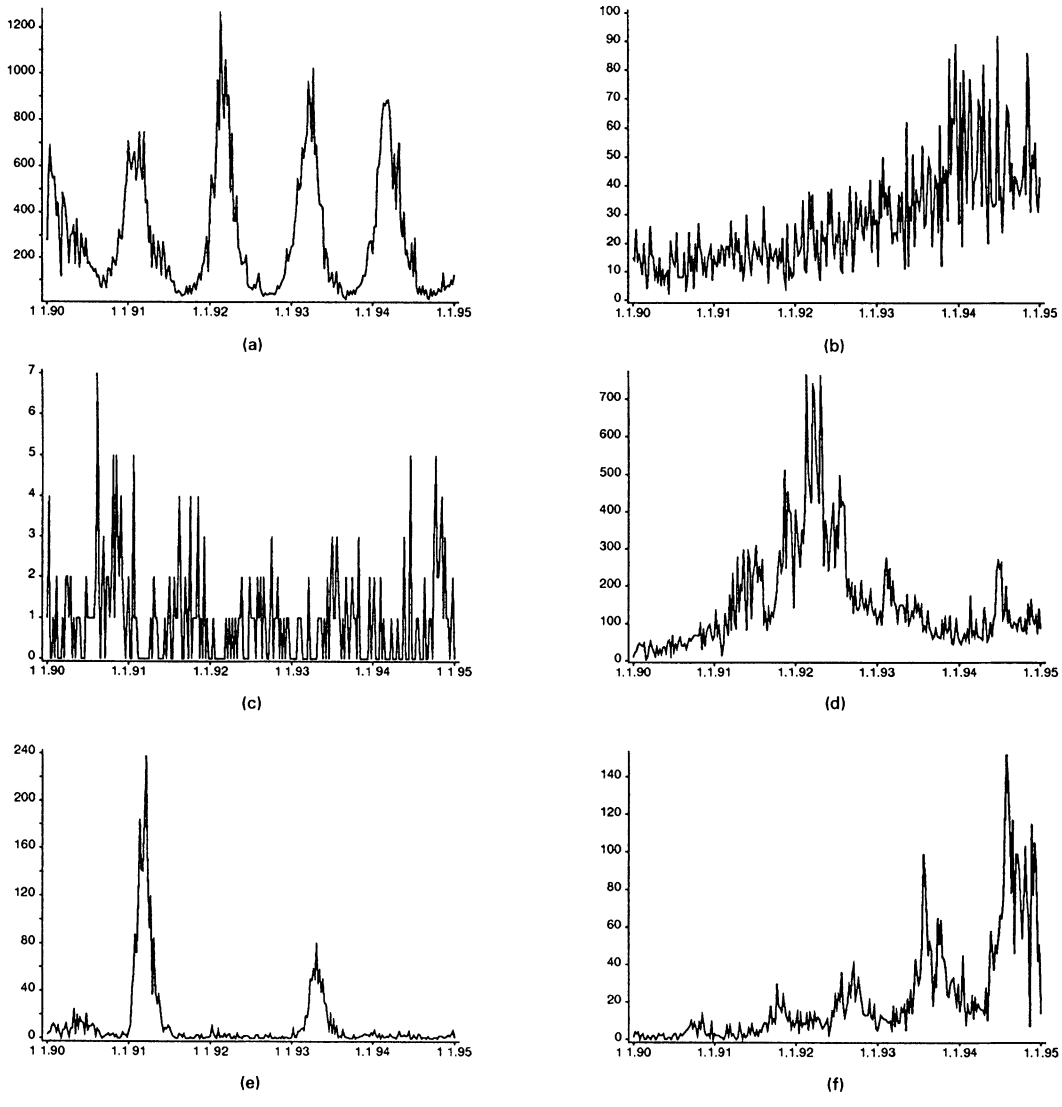

Fig. 1. Weekly count of selected organisms reported to the CDSC, 1990–95: (a) rotavirus; (b) *Clostridium difficile*; (c) *Salmonella derby*; (d) *Shigella sonnei*; (e) influenza B; (f) *Salmonella typhimurium* DT104

2.3. *Requirements of Routine Scanning System*

The main requirements of a routine scanning system are timeliness, sensitivity and specificity, together with readily interpretable outputs. Timeliness and sensitivity are required to ensure that outbreaks are detected in time for interventions to take place, but this should not be at the cost of a disproportionately high false positive detection rate, which would result in wasted time and effort, and undermine confidence in the system.

These requirements to a large extent determine the statistical features of the system. In practice the week's data must be processed automatically, typically over a

week-end, to leave time for epidemiological investigations. Thus statistical methods requiring careful model checking or resetting of parameters, such as time series and cumulative sums, are inappropriate since an organism-specific modelling strategy would be far too time consuming. Ideally, one single robust algorithm is required for all organisms. This must be sufficiently flexible to accommodate the wide variety of organism frequencies, seasonal patterns, underlying trends and extraneous noise in the data and illustrated in Fig. 1, without sacrificing sensitivity and specificity. Statistical methods relying heavily on stationarity or strong distributional assumptions, or requiring complex model selection procedures, must therefore be discounted.

After reviewing the available methodologies, it was decided to opt for a log-linear regression model, adjusted for overdispersion, seasonality, secular trends and past outbreaks. The model is used to calculate an expected value for the current week based on historical data, together with a threshold above which an observed count is declared to be unusual. The statistical algorithm, its validation and shortcomings are described in subsequent sections. Throughout, in keeping with the nature of the application, the approach is decidedly empirical, the emphasis being on achieving an algorithm which is suitably robust with respect to underlying assumptions.

3. STATISTICAL ISSUES

3.1. *Choice of Reference Date*

Delay inevitably occurs between the date at which an individual is infected and the date at which the report of the infection, complete with details of the organism involved, is received at the CDSC. In the intervening period a faecal, oral, serum or other specimen must be taken from the infected individual and analysed to determine the infectious agent. This process may take several days. The time series of organism counts by week of infection (or week of specimen collection, the closest available proxy since actual infection dates are rarely known) is thus inevitably subject to bias, the reported counts in recent weeks representing only a fraction of the true values since only cases with short delays in reporting are included. In consequence, a detection system based on this time series would result in outbreaks being missed. Nor is it realistic to attempt to adjust recent counts to offset the reporting delay bias, which would involve organism-specific correction factors of questionable reliability. Thus the time series of counts by date of infection is inappropriate for the detection of outbreaks unless delays in reporting are so short that the reporting delay bias can be ignored. This is not so for the data received at the CDSC. For salmonellas reported in 1994 for instance, the distribution of delays between the date at which the specimen was taken and the date of report was as follows: 0 weeks, 0.041; 1 week, 0.401; 2 weeks, 0.242; 3 weeks, 0.194; 4 weeks, 0.054; 5 weeks or more, 0.068. Instead, the system is based on the time series of organism counts by date of report to the CDSC. This has the added advantage that dates of report are always known, which is not so for specimen dates, the completeness of which varies considerably between organisms.

The use of the date of report as the reference date eliminates the major biases due to delays in reporting. Nevertheless, the distribution of the delay between infection and report influences the timeliness and sensitivity of the detection system by introducing additional variability. To investigate formally the effect of delays, let

$F(x)$ be the cumulative distribution of delays in reporting, x denoting the time between infection of the individual and organism report. Suppose that, at time t, $X(t)$ specimens are collected. Then the expected number of reports at time t is

$$Y(t) = \int_0^t X(u) f(t-u)\,\mathrm{d}u$$

where $f(x)$ is the density function of the reporting delay distribution. Suppose that $X(t) = 1$ for $t \in [0, \delta]$ and $X(t) = 0$ otherwise. Then given a threshold value $\alpha < 1$ which $Y(t)$ must reach for the outbreak to be detected, the detection time t_d satisfies

$$\alpha = F(t_\mathrm{d}) - F(t_\mathrm{d} - \delta)$$

where $F(t-\delta) = 0$ for $t < \delta$. Outbreaks of long duration δ will be detected at $t_\mathrm{d} = F^{-1}(\alpha)$, whereas short outbreaks will be detected close to the mode t_m of f, but only provided that it satisfies $f(t_\mathrm{m}) \geqslant \alpha/\delta$. Hence sensitivity to outbreaks of short duration is improved if the delay distribution is highly peaked. Thus, although it is highly desirable to reduce mean reporting delays to a minimum, it is also important to ensure that different reporting laboratories follow similar procedures, to reduce the variability of the delays. For most organisms reported to the CDSC, the delay distribution is highly peaked with mode at 1 week. Overall, the proportions of isolates reported within 4 weeks of the date at which the specimen was taken is in excess of 85%. Thus when an outbreak is detected it is unlikely to have started more than a few weeks earlier.

3.2. *Trends and Seasonality*

Trends are readily incorporated by fitting a linear time variable in the regression model. Seasonal effects are allowed for by basing the threshold calculation only on counts from comparable periods in past years. This is the same mechanism as that used by Stroup *et al.* (1989). These base-line weeks are specified by integers b and w, b denoting the number of years back and w the window half-width. If the current week is x of year y then only data for weeks from $x - w$ to $x + w$ of years from $y - b$ to $y - 1$ are used, giving a total $n = b(2w + 1)$ base-line weeks. The desirability of a large value of n to increase precision must be weighed against the need for a narrow window width in relation to the timescale of seasonal variations, whereas b is constrained by the need to base comparisons only on the recent past in view of changes in reporting procedures and laboratory methods. The values used are $b = 5$ and $w = 3$, giving a total of $n = 35$ base-line values. Since organism classifications are occasionally updated, for some the historical data do not extend for the full 5 years.

3.3. *Regression Model*

The base-line count y_i corresponding to the base-line week t_i is assumed to be distributed with mean μ_i and variance $\phi\mu_i$. Weekly base-line counts are assumed to be independent, an assumption which will be re-examined later. The only systematic effect to be included in the model is a linear time trend in the frequency of reports. Thus the systematic component of the model is

$$\log \mu_i = \alpha + \beta t_i,$$

where t_i is measured in weeks, typically taking the values 1–7, 53–59, etc. Estimates are obtained by a quasi-likelihood method. The dispersion parameter ϕ is estimated by

$$\hat{\phi} = \max\left\{\frac{1}{n-p}\sum_{i=1}^{n}\omega_i\frac{(y_i-\hat{\mu}_i)^2}{\hat{\mu}_i}, 1\right\},$$

where ω_i is a weight to be described later, and $p = 1$ or $p = 2$ depending on whether a time trend is fitted. Let t_0 denote the current week and y_0 the current organism count. The expected count is estimated by

$$\hat{\mu}_0 = \exp(\hat{\alpha} + \hat{\beta}t_0).$$

The linear time trend is included in the model only if the historical data span at least 3 years, if it is significant at the 5% level and if

$$\hat{\mu}_0 \leqslant \max\{y_i: i = 1 \ldots n\}.$$

This last condition protects against unrealistic extrapolations of the temporal trend. In practice, in a typical week, the time trend is included in the model for about a quarter of the organisms reported in that week.

3.4. *Distributional Assumptions*

Fig. 2 shows the scatterplot and running median (on 19 values) of the dispersion parameter $\hat{\phi}$ against the expected value $\hat{\mu}_0$ for 240 organisms reported during one week in 1995; these values have been adjusted to remove some of the effects of past outbreaks, as described later. For $\mu_0 < 1$, substantial overdispersion with respect to

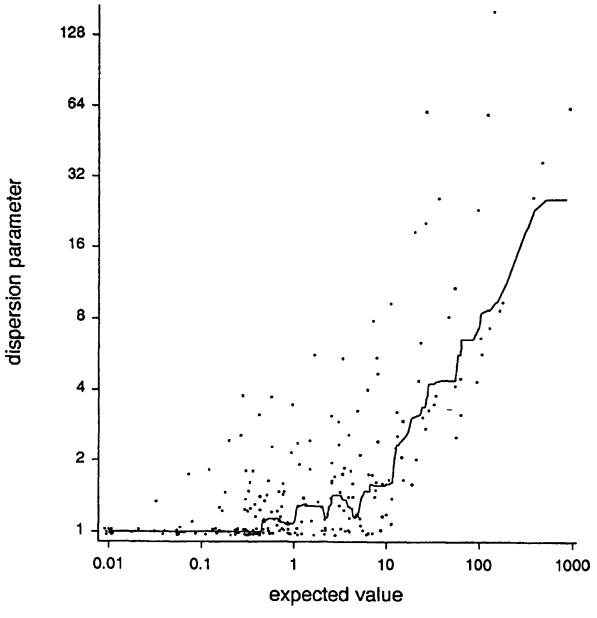

Fig. 2. Jittered scatterplot and running median on 19 values of dispersion parameter ϕ against expected value μ_0 for 240 organisms reported in week 14, 1995

the Poisson distribution is uncommon and the median value of ϕ is close to 1. For $\mu_0 > 10$, the median value of ϕ increases rapidly. This observation provides some rationale for the calculations in the next sections, which are based on methods for Poisson-distributed data, with the dispersion factor ϕ to allow for overdispersion. The underlying assumption is that counts of the less frequent organisms are typically Poisson distributed, whereas those of more frequent organisms may be assumed to be approximately normal. The validity of the methods is assessed by using simulated data; overdispersion is generated by assuming that the Poisson mean varies according to a gamma distribution with mean μ and variance $\mu(\phi - 1)$, thus resulting in negative binomial data with mean μ and variance $\phi\mu$.

3.5. *Threshold Calculation*

Organisms with low counts generally have highly skewed distributions. Some account of this needs to be taken in calculating the threshold value, to retain a broadly constant false positive probability over the wide range of values of μ_0. In this context the false positive probability is the probability of a count exceeding the threshold when there is no outbreak. A correction for skewness is achieved by applying the $\frac{2}{3}$-power transformation. Provided that the cumulants of the distribution of y_0 are $O(\mu_0)$, which is true of Poisson and negative binomial counts, then for large values of μ_0 Taylor series approximations yield

$$E(y_0^{2/3}) \doteq \mu_0^{2/3},$$

$$\text{var}(y_0^{2/3}) \doteq \frac{4}{9}\,\phi\mu_0^{1/3}$$

and

$$\text{var}(\hat{\mu}_0^{2/3}) \doteq \frac{4}{9}\,\mu_0^{-2/3}\,\text{var}\,(\hat{\mu}_0).$$

The prediction error variance on the $\frac{2}{3}$-power scale is

$$\text{var}(y_0^{2/3} - \hat{\mu}_0^{2/3}) \doteq \frac{4}{9}\,\tau\mu_0^{1/3},$$

where

$$\tau = \phi + \text{var}(\hat{\mu}_0)/\mu_0.$$

Hence the interval (L, U) with

$$U = \hat{\mu}_0\left\{1 + \frac{2}{3}\,z_\alpha\left(\frac{\hat{\tau}}{\hat{\mu}_0}\right)^{1/2}\right\}^{3/2},$$

$$L = \hat{\mu}_0 \max\left\{\left\{1 - \frac{2}{3}\,z_\alpha\left(\frac{\hat{\tau}}{\hat{\mu}_0}\right)^{1/2}\right\}^{3/2}, 0\right\}$$

defines an approximate $100(1 - 2\alpha)\%$ prediction interval for y_0, where z_α is the $100(1 - \alpha)$-percentile of the normal distribution. Organism counts outside this

interval are considered aberrant, those greater than the upper threshold U being flagged as possible outbreaks. For Poisson-distributed counts the $\frac{2}{3}$-power transformation induces approximate symmetry, thus resulting in more accurate thresholds. The thresholds for the more frequent organisms are largely unaffected by the transformation. Fig. 3 shows the false positive probability as a function of the mean organism frequency for Poisson and negative binomial counts, with $z_\alpha = 2.58$, corresponding to a nominal false positive probability $\alpha = 0.005$. Also shown are the results without transformation and with the square-root transformation (the variance stabilizing transformation for the Poisson distribution). These alternative definitions of the threshold are respectively

$$U = \hat{\mu}_0\left\{1 + z_\alpha\left(\frac{\hat{\tau}}{\hat{\mu}_0}\right)^{1/2}\right\},$$

$$U = \hat{\mu}_0\left\{1 + \frac{1}{2}\, z_\alpha\left(\frac{\hat{\tau}}{\hat{\mu}_0}\right)^{1/2}\right\}^2.$$

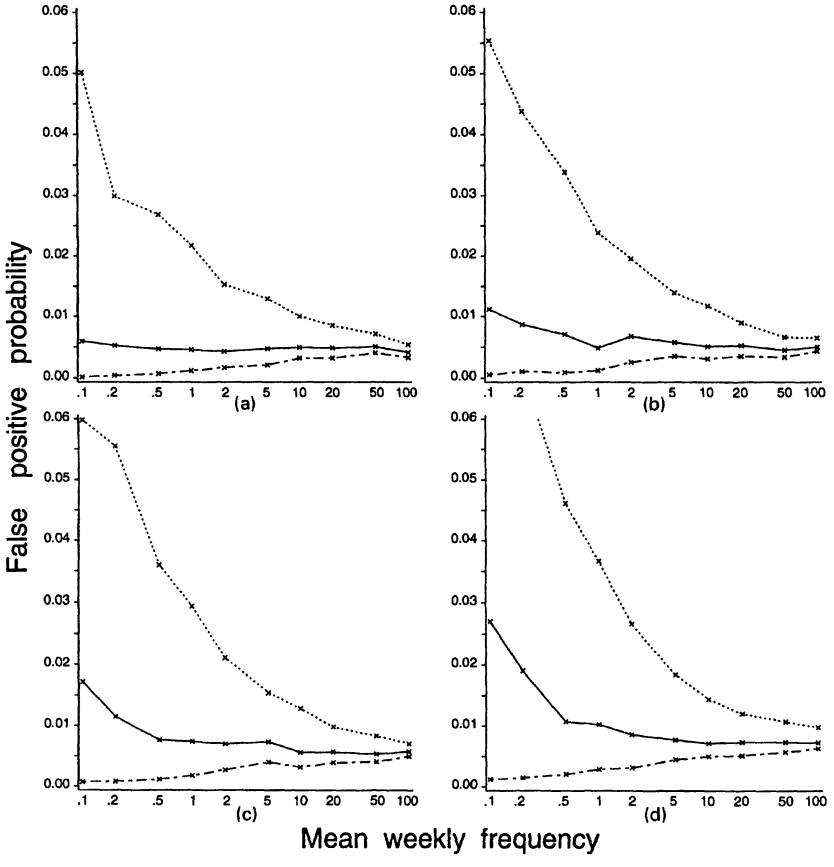

Fig. 3. False positive probability by mean weekly organism frequency (·····, no transformation; ———, $\frac{2}{3}$-power transformation; ·–·–·, square-root transformation): (a) Poisson; (b) negative binomial ($\phi = 2$); (c) negative binomial ($\phi = 3$); (d) negative binomial ($\phi = 5$)

Fig. 3 shows that, for Poisson data, the threshold based on the $\frac{2}{3}$-power transformation results in a false positive rate which remains approximately constant in the range of interest, namely $\mu_0 = 0.1$–100. For negative binomial data the false positive rate is roughly constant for $\mu_0 > 0.5$. For fixed μ_0, the false positive rate generally increases slightly with ϕ.

3.6. *Correcting for Past Outbreaks*

A major difficulty with the regression approach is how to reduce the influence of base-line counts in weeks coinciding with past outbreaks. The inclusion of past outbreak values in the threshold calculations results in thresholds which are too high and hence reduces sensitivity. Manual monitoring of the base-lines to identify outliers and to exclude them from the calculations is impractical. Instead, a reweighting procedure is used to reduce the influence of high base-line counts. Using initial estimates of μ_i and the dispersion parameter ϕ obtained with $\omega_i = 1$, residuals are defined as follows:

$$s_i = \frac{3}{2\hat{\phi}^{1/2}} \frac{y_i^{2/3} - \hat{\mu}_i^{2/3}}{\hat{\mu}_i^{1/6}(1 - h_{ii})^{1/2}},$$

where h_{ii} are the diagonal elements of the hat matrix. For Poisson data, for which $\phi = 1$, the s_i are the standardized Anscombe residuals (Davison and Snell, 1991). The weights are then defined by

$$\omega_i = \begin{cases} \gamma s_i^{-2} & \text{if } s_i < 1, \\ \gamma & \text{otherwise,} \end{cases}$$

where γ is a constant such that $\Sigma \omega_i = n$. The weighting function was chosen on empirical grounds to assign very low weights to counts with large residuals. Table 2 shows the results of simulations with Poisson data to investigate the effect of reweighting on the threshold values, both with and without outbreaks in the base-lines. Such outbreaks were simulated by adding a constant, representing the size of

TABLE 2

Median upper threshold values with and without reweighting by organism frequency, outbreak standard deviations (on the $\frac{2}{3}$-power scale) and number of base-lines affected

No. of outbreak base-lines	*Reweighting*	*Median upper threshold values for the following underlying mean frequencies:* 1		10		100	
		Outbreak standard deviations 3	5	3	5	3	5
0	No	4.68	4.68	19.7	19.7	128.3	128.3
	Yes	4.24	4.24	18.8	18.8	125.9	125.9
1	No	5.39	7.55	21.1	24.2	131.5	138.6
	Yes	4.46	4.84	19.2	19.5	126.9	127.6
2	No	6.16	9.68	22.9	28.2	135.3	147.7
	Yes	4.75	5.69	19.6	20.7	127.8	129.5
5	No	7.81	13.7	26.9	36.6	145.3	168.8
	Yes	6.08	9.37	22.2	26.3	133.1	142.0

outbreak, to some of the base-line counts. The constants chosen correspond to multiples of the standard deviation on the $\frac{2}{3}$-power scale, to maintain comparability for different organism frequencies. The relationship between the number of standard deviations, z, on the $\frac{2}{3}$-power scale, and the corresponding size of outbreak $\delta(z)$ measured in number of organisms is

$$\{\mu_0 + \delta(z)\}^{2/3} - \mu_0^{2/3} \doteq \frac{2}{3} z\phi^{1/2}\mu_0^{1/6},$$

whence

$$\delta(z) \doteq \mu_0\left[\left\{1 + \frac{2}{3} z\left(\frac{\phi}{\mu_0}\right)^{1/2}\right\}^{3/2} - 1\right].$$

Median threshold values calculated with the reweighting are little affected when there is no outbreak in the base-lines; thus the reweighting does not materially reduce the specificity of the system. In contrast the reweighting substantially reduces the bias in the threshold value induced by outbreaks in the base-lines. For instance, for an organism with an underlying mean weekly frequency of 10, the median threshold value is 19.7 without and 18.8 with reweighting. A departure of five standard deviations above the mean on the $\frac{2}{3}$-power scale corresponds to an additional count of about 19 organism reports, as would occur in an outbreak of size 19. If this value is added to five of the 35 base-line counts, the median threshold value rises to 36.6. The reweighting reduces this to 26.3. The reweighting thus substantially alleviates the effect of past outbreaks but does not eliminate it.

3.7. *Correlations between Base-line Counts*

The threshold calculation ignores serial correlations between base-line counts. This effect has been investigated by Kafadar and Stroup (1992). For infrequent organisms occurring sporadically no appreciable correlations would be expected. For more frequent organisms, however, the dynamics of organism transmission and residual effects of seasonality might be expected to introduce some correlation between base-line counts within each year. The effect of such correlations is investigated by assuming that

$$\operatorname{corr}(y_i, y_j) = \begin{cases} \rho & \text{if } t_i, t_j \in \text{same year,} \\ o(\rho) & \text{otherwise.} \end{cases}$$

For simplicity we assume that there are no secular time effects. Then the predicted count at time t is $\hat{\mu}_0 = \Sigma\, y_i/n$ and

$$\operatorname{var}(\hat{\mu}_0) = \frac{\phi\mu_0}{n}\{1 + 2w\rho + o(\rho)\}.$$

Thus the quantity τ in the prediction error variance, which is $\phi(1 + 1/n)$ when $\rho = 0$, becomes

$$\tau^* = \phi\left[1 + \frac{1}{n}\left\{1 + 2w\rho + o(\rho)\right\}\right].$$

This increases the threshold by

$$U^* - U = \mu_0 \left[\left\{ 1 + \frac{2}{3} z_\alpha \left(\frac{\tau^*}{\mu_0} \right)^{1/2} \right\}^{3/2} - \left\{ 1 + \frac{2}{3} z_\alpha \left(\frac{\tau}{\mu_0} \right)^{1/2} \right\}^{3/2} \right],$$

which for large μ_0 and n is approximately

$$U^* - U \doteq z_\alpha (\phi \mu_0)^{1/2} \frac{w\rho}{n}.$$

Hence correlation between counts in adjacent weeks introduces some bias in the detection threshold U. However, this bias is generally small: for instance, for $n = 35$, $w = 3$, $z_\alpha = 3$, $\mu = 100$, $\phi = 9$ and $\rho = 0.5$, $U^* - U$ is less than 5. Thus for frequent organisms the effect of serial correlation has little effect on sensitivity or specificity. For infrequent organisms we would expect $\rho = 0$ in non-epidemic conditions, namely those under which the thresholds are computed. Serial correlation is therefore ignored in the threshold calculation.

3.8. *The Algorithm*

Each week, tables of organism counts are updated by using a classification level that is appropriate for epidemiological analysis; for instance salmonellas are classified into serotypes and phage types. The following algorithm is then applied to the vector of counts of each organism reported. An initial model is fitted and initial estimates $\hat{\mu}_i$ and $\hat{\phi}$ are obtained. Weights are calculated as described above and the model is refitted. A revised estimate of ϕ is obtained and the model is rescaled. If not significant, the trend is omitted and the procedure repeated. Finally the threshold value is calculated. The output consists of a listing of organisms in order of exceedance score, defined as

$$X = \frac{y_0 - \hat{\mu}_0}{U - \hat{\mu}_0}.$$

The exceedance score is set to 0 if fewer than five reports were received in the past 4 weeks. Organisms with $X > 1$ are then flagged for more detailed investigation. The minimum outbreak size of 5 over 4 weeks reduces the likelihood that sporadic cases of infrequent organisms are flagged. The choice of threshold parameter z_α is determined empirically to keep the number of detections to manageable levels. The value 2.58, formally corresponding to a 99% prediction interval (L, U) and hence a false positive probability of 0.005, has been adopted and typically results in about 20 organisms being flagged each week.

To assist in the interpretation of the exceedance scores, a set of graphs is produced for each organism, showing the distribution of organism counts by week, region and age group. An example is given in Fig. 4. The algorithm and the graphs are generated using GLIM4 (Francis *et al.*, 1993), the whole system being triggered automatically each week-end to aggregate the week's reports and to produce the output for the following Monday. The list of organisms flagged by the system is reviewed by epidemiologists at the CDSC. A small number of organisms deemed worthy of further investigation are then identified, this final selection being based on subject-matter knowledge.

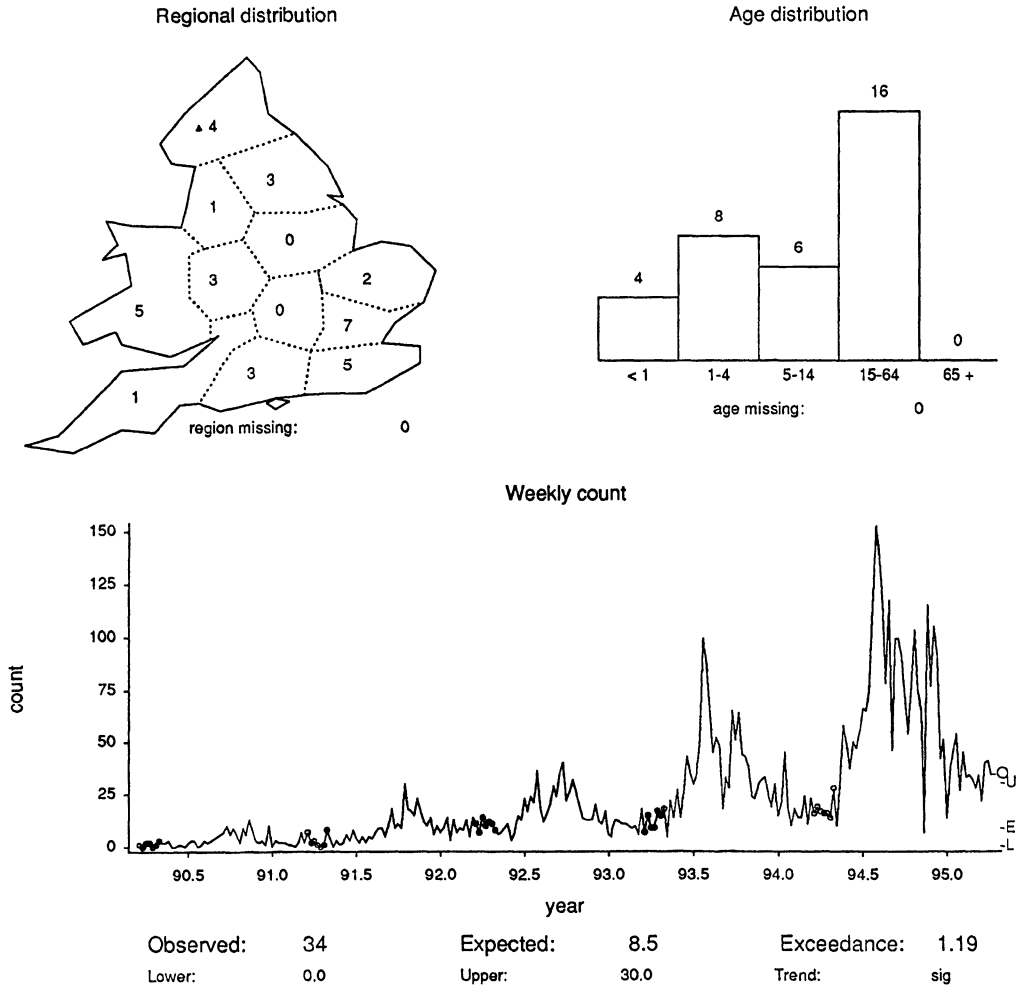

Fig. 4. Output screen for *Salmonella typhimurium* DT104: the triangle against the Northern region count indicates an unusually high value

4. EVALUATION

The primary aim of the system is to assist in the identification of country-wide outbreaks or changes in trend, and a limited evaluation of an earlier version may be found in Farrington and Beale (1993). The performance of the statistical algorithm described in the present paper was assessed by simulation. In addition, the system's ability to flag unusual events of national importance was evaluated by reviewing reports of national incidents for 1994. Finally, the operational performance of the system was investigated by means of a case-study.

4.1. *Simulation*

The algorithm as a whole was evaluated by using simulated Poisson and negative binomial data. From an operational point of view it is important to know what size of outbreak may be detected for any given organism. The probability of detecting an

outbreak was estimated for organism frequencies of 0.1, 1, 10 and 100 per week, assuming that no outbreaks occurred during the base-line weeks, for a range of current outbreak sizes. These outbreak values were added to the randomly generated Poisson or negative binomial count for the current week. The outbreak is detected if the current total is greater than or equal to the threshold calculated from the 35 base-line values. Results are shown in Fig. 5. For organisms with ϕ no greater than 2, which includes most organisms with mean weekly frequency below 10, the detection probability rises rapidly with outbreak size. For Poisson-distributed counts with a mean weekly count of 0.1, 80% detection probability is reached for outbreaks of size 2; for mean weekly counts of 1, it occurs for outbreaks of size 4–5. For more common organisms with substantial overdispersion, larger outbreaks are required to trigger the system in view of the increased background variability. Thus, for organism counts with weekly mean 10 and $\phi = 2$, 80% detection probability occurs for outbreaks of size 15–16, whereas for organisms with weekly mean 100 and $\phi = 10$ this rises to 104. The outbreak sizes quoted above for a detection probability of 80% all correspond to departures of approximately three standard deviations on the $\frac{2}{3}$-power scale. For a given outbreak size, the detection probability will be reduced if base-line values coincide with past outbreaks.

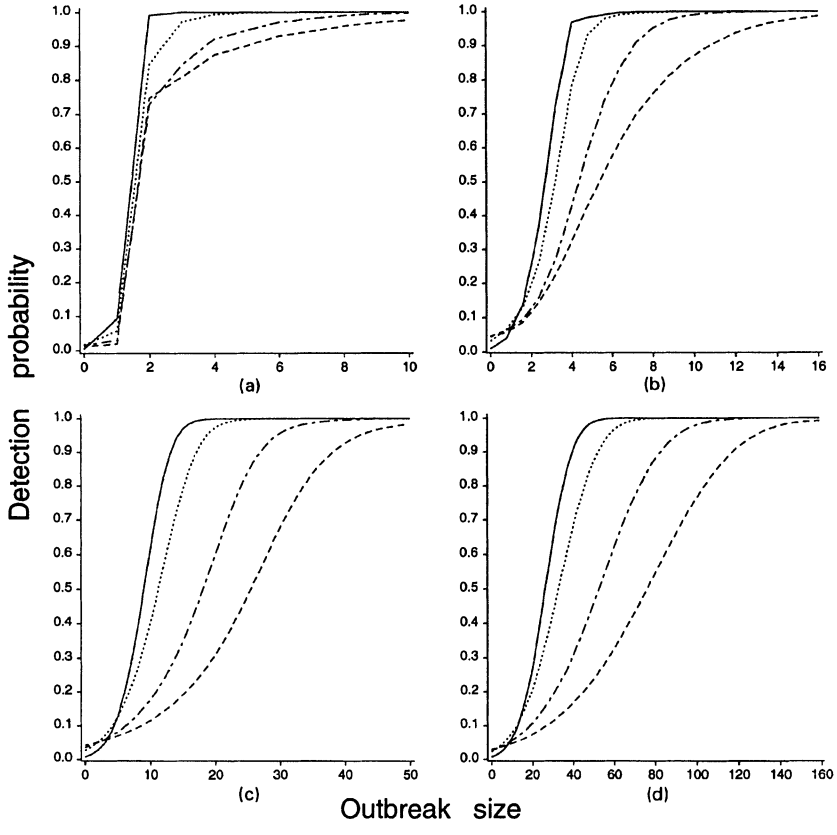

Fig. 5. Probability of detection by outbreak size—simulation results based on Poisson data (———) and negative binomial data with $\phi = 2$ (·····), $\phi = 5$ (·–·–·) and $\phi = 10$ (- - - -): (a) mean 0.1; (b) mean 1; (c) mean 10; (d) mean 100

4.2. *Review of National Incidents*

Events of national importance were documented by reviewing the 1994 volume of the *Communicable Disease Report* and from records of national incidents investigated in 1994. A total of 11 such events were identified. In three instances the incidents were detected through surveillance systems not based on laboratory reports. The remaining eight incidents included increased laboratory reports of *Salmonella javiana*, *Cryptosporidium*, *Enterobacter cloacae*, *Salmonella virchow* phage type 26, *Legionella pneumophilia* group 1, *Staphylococcus aureus*, *Salmonella bovismorbificans* and *Shigella sonnei*. All except two were flagged by the system. The two failures were *Cryptosporidium* and *Shigella sonnei*, in the first case because the incident involved several small local outbreaks which did not amount to a large increase in national reports, and in the second due to the system's inability to adjust for the very high base-line counts (see Fig. 1).

4.3. *Case-study*

In early February 1995, a cluster of cases of *Salmonella agona* was reported to the CDSC. Epidemiological and microbiological investigations revealed that the outbreak of this relatively unusual salmonella serotype was caused by contamination of a popular kosher snack. It soon became apparent that the outbreak was not limited to Britain, and that worldwide distribution of this food-stuff had resulted in cases in other countries. Although the detection system was in operation at the time, it did not trigger the investigation; this was initiated by more traditional observational surveillance methods. The performance of the system with regard to *Salmonella agona* was therefore reviewed retrospectively.

Fig. 6 shows the time series of reports from 1990 to week 20, 1995. Also shown are a more detailed graph of weekly reports from week 33 of 1994 (beginning August 15th) to week 20 of 1995 (beginning May 15th) along with the exceedance scores for *Salmonella agona* calculated in each of these weeks. The exceedance score first exceeded 1 in week 44 of 1994 (beginning October 31st), in which eight isolates were reported. However, this was an isolated peak, the exceedance score returning to a very low value in the following week. This excess could reasonably have been attributed to chance or some spurious reporting effect. The exceedance score next exceeded 1 in weeks 50 and 51 of 1994 (beginning December 12th and 19th). This was followed by a trough over the next 3 weeks, corresponding to the Christmas and New Year break. The exceedance score again rose above 1 in week 3 of 1995 (beginning January 16th) and remained above 1 for 4 weeks.

The outbreak was therefore repeatedly flagged by the system as early as mid-December 1994. In spite of the unambiguous message revealed by the exceedance scores, the observed excesses did not prompt further investigations until it became apparent that the outbreak was restricted to a well-defined population. This was due to the unexceptional appearance of the time series of weekly reports during this period, the high exceedance scores being incorrectly dismissed as false positives.

5. DISCUSSION

The system described in this paper is intended as an aid to the detection of outbreaks to supplement other more intensive surveillance methods by routinely

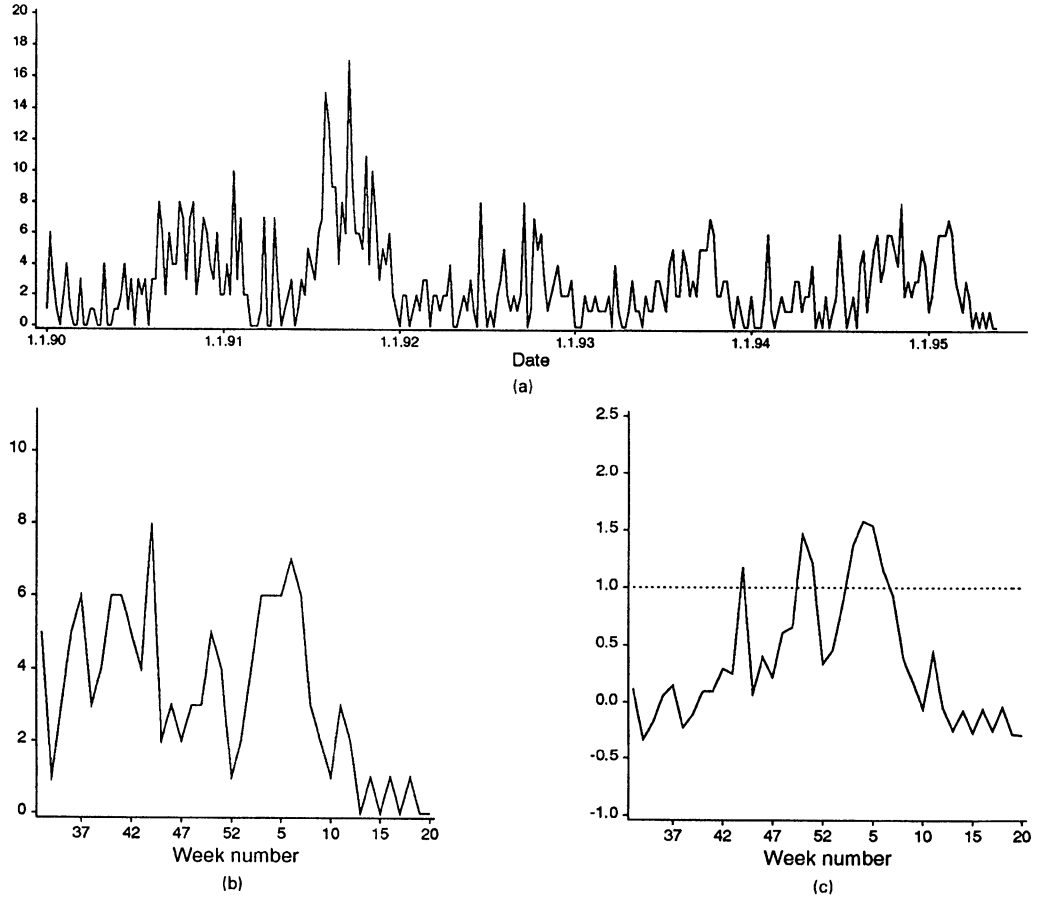

Fig. 6. Weekly counts and exceedance scores for *Salmonella agona*: (a) weekly reports, 1990–95; (b) reports between week 33, 1994, and week 20, 1995; (c) exceedances between week 33, 1994, and week 20, 1995

scanning all organism reports received at the CDSC. Its strength lies in its being automated, providing a timely short list of potential problems. For uncommon organisms the system is highly sensitive and will detect small increases in weekly counts with high probability. For more common organisms, especially those which exhibit substantial variability, only large excesses will be detected. In particular, small localized outbreaks of common organisms are unlikely to be identified. However, such outbreaks are usually noticed at the local level anyway. The main purpose of the system is to detect national outbreaks, especially those which may not register a substantial increase in reports in any particular region.

Experience with the system suggests that roughly 40% of the excesses identified correspond to outbreaks or other events of public health interest. A further 30% correspond to increases of questionable epidemiological interest, for instance known long-term upward trends, whether genuine or due to increased testing, or chance occurrences of sporadic cases. The remaining 30% are due to idiosyncracies in reporting, such as batching of reports by one laboratory. In respect of the last

category, the system provides a monitoring and auditing tool which, if used appropriately, can generate improvements to surveillance procedures. A weakness of the system is its insensitivity when the base-line values on which the threshold calculation is based coincide with past outbreaks. The reweighting procedure described alleviates the problem but cannot eliminate it. An alternative approach, in which all counts which were flagged as exceptional when they were current are excluded from base-lines, is being investigated.

The case-study of *Salmonella agona* highlights both the potential and the shortcomings of automated detection procedures. In this instance, the system detected the outbreak several weeks before it was identified by more traditional epidemiological methods. However, the warnings went unheeded owing to a lack of a clear peak in reports. This emphasizes the need for further close collaboration between statisticians and epidemiologists in formalizing the procedures employed to review the system's output.

REFERENCES

Chen, R., Connelly, R. R. and Mantel, N. (1993) Analysing post-alarm data in a monitoring system in order to accept or reject the alarm. *Statist. Med.*, **12**, 1807–1812.

Davison, A. C. and Snell, E. J. (1991) Residuals and diagnostics. In *Statistical Theory and Modelling* (eds D. V. Hinkley, N. Reid and E. J. Snell). London: Chapman and Hall.

Ederer, F., Myers, M. H. and Mantel, N. (1964) A statistical problem in space and time: do leukemia cases come in clusters? *Biometrics*, **20**, 626–638.

Farrington, C. P. and Beale, A. D. (1993) Computer aided detection of temporal clusters of organisms reported to CDSC. *Communcbl. Dis. Rep. Rev.*, **3**, R78–R82.

Francis, B., Green, M. and Payne, C. (1993) *The GLIM System: Release 4 Manual.* Oxford: Oxford University Press.

Kafadar, K. and Stroup, D. F. (1992) Analysis of aberrations in public health surveillance data: estimating variances on correlated samples. *Statist. Med.*, **11**, 1551–1568.

Langmuir, A. D. (1963) The surveillance of communicable diseases of national importance. *New Engl. J. Med.*, **268**, 182–192.

Naus, J. I. (1965) The distribution of the size of the maximum cluster of points on a line. *J. Am. Statist. Ass.*, **60**, 532–538.

Nobre, F. F. and Stroup, D. F. (1994) A monitoring system to detect changes in public health surveillance data. *Int. J. Epidem.*, **23**, 408–418.

Page, E. S. (1954) Continuous inspection schemes. *Biometrika*, **41**, 100–115.

Parker, R. A. (1989) Analysis of surveillance data with Poisson regression: a case study. *Statist. Med.*, **8**, 285–294.

Roger, J. H. (1977) A significance test for cyclic trends in incidence data. *Biometrika*, **64**, 152–155.

Serfling, R. E. (1963) Methods for current statistical analysis of excess pneumonia–influenza deaths. *Publ. Hlth Rep.*, **78**, 494–506.

Stroup, D. F., Williamson, G. D. and Herndon, J. L. (1989) Detection of aberrations in the occurrence of notifiable diseases surveillance data. *Statist. Med.*, **8**, 323–329.

Tango, T. (1984) The detection of disease clustering in time. *Biometrics*, **40**, 15–26.

Tillett, H. E. and Spencer, I.-L. (1982) Influenza surveillance in England and Wales using routine statistics. *J. Hyg. Camb.*, **88**, 83–94.

Wallenstein, S. (1980) A test for detection of clustering over time. *Am. J. Epidem.*, **111**, 367–372.

Watier, L. and Richardson, S. (1991) A time series construction of an alert threshold with application to *S. bovismorbificans* in France. *Statist. Med.*, **10**, 1493–1509.

Weatherall, J. A. C. and Haskey, J. C. (1976) Surveillance of malformations. *Br. Med. Bull.*, **32**, 39–44.