Epidemiol. Infect. (1999), **122**, 103–110. Printed in the United Kingdom

Automated outbreak detection: a quantitative retrospective analysis

L. STERN[1]* AND D. LIGHTFOOT[2]

[1] *Department of Computer Science and* [2] *Microbiological Diagnostic Unit, The University of Melbourne, Parkville, Victoria 3078, Australia*

(*Accepted 18 August 1998*)

SUMMARY

An automated early warning system has been developed and used for detecting clusters of human infection with enteric pathogens. The method used requires no specific disease modelling, and has the potential for extension to other epidemiological applications. A compound smoothing technique is used to determine baseline 'normal' incidence of disease from past data, and a warning threshold for current data is produced by combining a statistically determined increment from the baseline with a fixed minimum threshold. A retrospective study of salmonella infections over 3 years has been conducted. Over this period, the automated system achieved $> 90\%$ sensitivity, with a positive predictive value consistently $> 50\%$, demonstrating the effectiveness of the combination of statistical and heuristic methods for cluster detection. We suggest that quantitative measurements are of considerable utility in evaluating the performance of such systems.

INTRODUCTION

As computer technology has permeated hospital and public health laboratories, an increasing number of epidemiological databases are being maintained. Particularly where maintained on a national or international scale, these databases contain a wealth of information and have the potential to inform us about many aspects of disease [1, 2]. Because the relevant parameters are not always known when the data are collected, the general approach is to collect and store as much data as possible. While these databases are a rich source of multi-dimensional data, in many cases their potential is under-realized.

On an international level, food-borne illness has increased in recent years, rather than decreased [3–5]. While some of the increase can be attributed to improved reporting practices, changes in agricultural and food manufacture methods and a changing population have also led to increased disease incidence [3]. There is a need for a means for rapid detection of point-source outbreaks of food-borne gastroenteritis. This would allow the timely introduction of containment measures to reduce the extent of the outbreak and its attendant human illness and financial loss [6].

We have a comprehensive notification scheme of enteric pathogens in Australia, the National Enteric Pathogens Surveillance Scheme (NEPSS) (formerly the National Salmonella Surveillance Scheme), which has been developed and maintained at the Microbiological Diagnostic Unit of The University of Melbourne since 1980. The associated database contains data on isolates from most cases of gastroenteritis in Australia caused by *Salmonella enterica* and *Shigella* species, as well as data on isolates of a number of different bacterial species. The bulk of the data involves *Salmonella enterica*, which is divided into > 2300 groups (serovars) on the basis of serology.

* Author for correspondence.

https://doi.org/10.1017/S0950268898001939 Published online by Cambridge University Press

All isolates in the database have been serotyped and, where appropriate, most have been phage typed prior to data entry.

Many incipient outbreaks of salmonellosis have been detected by the NEPSS using an informal method, that is, unusual clusters of a particular serovar or serovar-phage type combination (strain) have been noticed by alert staff. The informal method relies on the fact that NEPSS receives reports from all over Australia; the concentration of reports in one place means that abnormal patterns are often recognized. In the absence of retrospective analysis, the precise rate of detection using this method was not known. We hypothesized that the informal method had a reasonable success rate for detecting outbreaks due to the more unusual strains of salmonella, but that it would be less effective in picking up single-source outbreaks of the more common strains, since local increases over a high background would be less likely to stand out. We designed an automated early warning system to see if outbreak detection could be improved, both with respect to the number of outbreaks detected and to the timeliness of detection. It flags clusters of a single serovar (and phage type, where appropriate) that are abnormal for the geographic location and time of year in which they occur, and which might warrant further epidemiological investigation. It also supports the initial stages of further investigation, by displaying additional information on demand.

We ran a prototype implementation of our early warning system, the Salmonella Potential Outbreak Targeting System (SPOT), version 1.0, in blind parallel with the informal system, and found that the automated system picked up more outbreaks than the informal system [7]. We have since improved the robustness of the automated system by introducing a hybrid statistic/heuristic algorithm to detect incipient outbreaks, in place of the purely heuristic approach of the prototype version. Both accuracy and efficiency have been improved by using a smoothing technique to deal with outlying data in the baseline, in place of the semi-manual monitoring used in the prototype. We have also expanded the system to include surveillance of *Shigella* species. In this paper we describe a 3-year retrospective analysis of salmonella surveillance using the improved version of the automated system, SPOT v2.0. We have measured the sensitivity of outbreak detection using the automated system and positive predictive value of the reports generated, and have compared these with results obtained using the informal, non-automated system. We discuss our results in light of the published results of other automated surveillance systems [8–11], and advocate the use of quantitative measures for comparing the effectiveness of different systems.

METHODS

System design

The early warning system works in two distinct phases. In phase 1, baseline values of expected occurrences are calculated, based on the previous 5 years' data. A baseline value is calculated for each serovar (or serovar and phage type), in each geographic location, and for each calendar month, and is stored for use in phase 2.

In phase 2, a warning threshold is calculated from the stored baseline values, and any incidence above the corresponding threshold is flagged. The length of time to be surveyed can be nominated by the user. Within the nominated time frame, isolations are always grouped by week, to avoid diluting – and consequently missing – a short-lived increase by averaging over a longer time period. The user can nominate the geographic area, which can be as large as all of Australia, or a smaller region. Where nationwide surveillance is nominated, the default is that isolations are grouped by state, in recognition of different patterns of distribution of different serovars and to avoid missing a relatively local outbreak by averaging over the entire country. The system can also be set to screen for occurrences of a single serotype, for repeated surveillance during an ongoing outbreak or suspected outbreak. The default configuration screens all of Australia for all *Salmonella* serotypes over the past 8 weeks.

National Enteric Pathogens Surveillance Scheme data

Data used were the NEPSS data for the years 1993–5. Data are recorded on a per isolation basis and stored using a commercial database (Ingres v6.5/05). For the purposes of this study, the reference date is the specimen date. Repeat isolations from the same person and isolations where the patient has recently been overseas are not included in the count of current isolations.

Baseline calculation

The baseline used in this paper is derived from the previous 5 years' data, smoothed to minimize the

https://doi.org/10.1017/S0950268898001939 Published online by Cambridge University Press

contribution of outlying data. Smoothing is accomplished using the compound smoothing technique 4253H [12], as implemented in the Minitab Statistical Software, release 9.1 [13].

Smoothing is performed as follows, for each serovar, phage type and geographic area. First, isolation counts are grouped by calendar month and year, resulting in 60 raw data points (5 years $\times$ 12 months). These raw data are then smoothed by a sequence of successive passes in which each point is replaced by a point that incorporates information from the neighbouring points in time. The sequence used calculates successive medians of 4, 2, 5 and 3 neighbouring data points, and finally the Hanning running average (H), a weighted running average where the data point for time t, $d(t)$, is replaced by $\frac{1}{4}d(t-1)+\frac{1}{2}d(t)+\frac{1}{4}d(t+1)$. The data are then reroughed, or polished, by applying the same sequence of smoothers (4253H) to the residuals (raw data minus smoothed values), and adding the smoothed residuals back to the smoothed data points. The differences between the smoothed value and the raw value for each data point are used to compute the standard deviation. The baseline is calculated by collapsing the smoothed data points into a yearly cycle, and taking the median of the five different smoothed values for each calendar month. Examination of several different baselines calculated using this method confirmed that the baseline retained seasonal changes and was an improvement over the method used in the prototype (data not shown).

A value is calculated for each month, for every combination of serovar, phage type and geographic area. Baseline values and standard deviations are recalculated annually and stored in a look-up table for fast retrieval during outbreak surveillance. This is sufficiently frequent to keep up with shifting trends, and at the same time sufficiently infrequent to minimize computation time. With >700 known *Salmonella* species in Australia, the look-up table is potentially very large, but considerable space is saved by storing only the non-zero values in this sparse distribution.

Outbreak detection

To detect outbreaks, current isolations over the nominated time period are grouped, and the counts for each week are compared to a warning threshold. The warning threshold is calculated at outbreak detection time, and has a statistical component and a heuristic component. The statistical component is calculated by taking the baseline value for the serovar, region and month in question, adding a multiple of the standard deviation, and adjusting this monthly cut-off to a weekly value. A heuristic component is an empirically derived value that allows us to filter out weeks where the isolations are too few to be worth following up, even though they might pass the statistical threshold. It is a fixed low number, indicating the largest number of occurrences in 1 week that will not appear in the report. Whenever the count of current isolations is greater than the statistical component and greater than the heuristic cut-off, it is flagged and appears in the generated report. For computational efficiency, the heuristic threshold is tested first.

Evaluation

The data presented in this paper were obtained by running the automated system over the data for all of Australia during 1993–5, one year at a time. Warning thresholds were calculated as

$$\text{Baseline} + 2{\cdot}0 \times \text{Standard Deviation},$$

and the heuristic cut-off factor was 2. These values were determined from pilot studies where both factors were varied.

'Sensitivity' and 'positive predictive value' are as described in the CDC *Guidelines for evaluating surveillance systems* [14]. 'Sensitivity' is defined as: 'relevant events noted/total relevant events', and is measured in terms of clusters salmonella retrieved. For these purposes, a cluster is defined as an unbounded number of contiguous weeks in which incidence of the strain in question is higher than the warning threshold, and also encompasses isolated weeks of high incidence at the start and end of a cluster, providing there are no more than 2 intervening weeks of normal or below-normal incidence between the main part of the cluster and the isolated week.

The 'positive predictive value' is defined as: 'relevant events noted/total events noted' and is measured in terms of rows in the report generated by the system.

Relevance for both measures was determined by the judgement of two epidemiologists as follows. Data for the test year and the 5 previous years was appropriately grouped and sorted, and the epidemio-

https://doi.org/10.1017/S0950268898001939 Published online by Cambridge University Press

logists were asked to score clusters in the test year that were sufficiently suspicious that further investigation would be desirable. These clusters were scored as relevant events, or 'potential outbreaks', regardless of whether evidence of epidemiological linkage was available.

Comparison data for assessing the 'informal' method came from the book of 'Suspected Outbreaks' kept by the NEPSS during the years 1993–5.

RESULTS

Surveillance report

The output from a typical SPOT report for a single state is shown in Table 1. This report shows all strains of salmonella for which the number of occurrences was above the warning threshold for this state during the 8-week period between 21 April and 15 June 1996. The nine rows in the report are grouped into five clusters. Clusters containing more than one row, such as those for *S*. Aberdeen and *S*. Mbandaka are highly suspicious as potential outbreaks and would be investigated further. In contrast, isolated clusters such as those for *S*. Birkenhead, *S*. Muenchen and *S*. Virchow might or might not be followed up by the epidemiologist, depending on a variety of factors. The SPOT application supports initial screening of these factors, by displaying postcode and patient name information for any row, on demand.

Sensitivity

Sensitivity is a measure of how well the system finds relevant events. Sensitivity obtained with the automated system over the 3-year period of this study is shown in Table 2, along with comparison values for sensitivity using the informal system. In both cases, the relevance standard is the retrospective evaluation of an epidemiologist.

As seen in Table 2, sensitivity is consistently higher than 90% when the automated system, SPOT v2.0, is used. This is considerably higher than the rate of detection of the informal method (59–69%).

Of particular interest are two large peaks detected by the automated system but not by the informal system. In one state there were > 100 isolations of *S*. Virchow over a 3-month period (Fig. 1*a*), with 30 in 1 week. In another state there were 45 isolations of *S*. Typhimurium, phage type 44 (PT 44) over a 3-month

Table 1. *Typical SPOT report for a single state, covering 8 weeks*

Organism	Week	Occurrences/ week	Expected occurrences
S. Aberdeen	13/05/96	4	1·15
S. Aberdeen	20/05/96	4	1·15
S. Birkenhead	27/05/96	3	1·51
S. Mbandaka	22/04/96	3	0·00
S. Mbandaka	06/05/96	4	0·00
S. Mbandaka	20/05/96	4	0·00
S. Mbandaka	27/05/96	3	0·00
S. Muenchen	13/05/96	3	0·82
S. Virchow	20/05/96	11	5·79

Table 2. *Sensitivity for SPOT v2.0 and informal system*

	Total potential outbreaks	Informal system		SPOT v2.0	
		Detected	Sensitivity (%)	Detected	Sensitivity (%)
1993	32	19	59	30	94
1994	57	39	68	52	91
1995	45	31	69	42	93

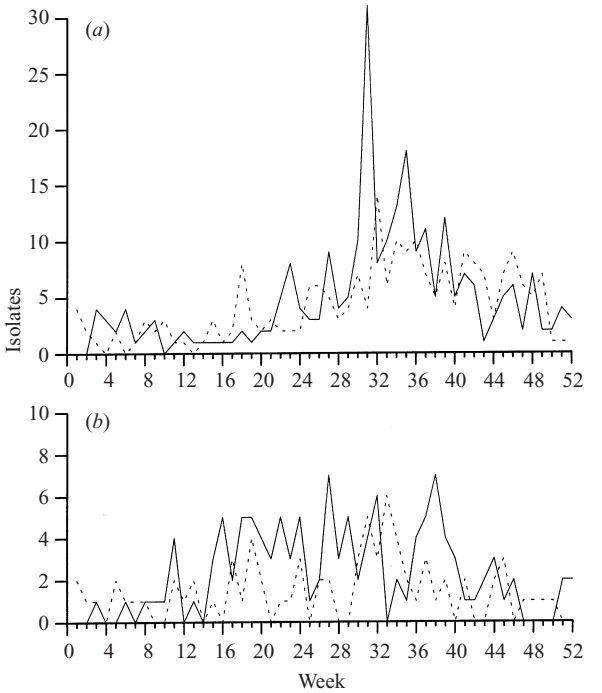

Fig. 1. Increased incidence of two serovars detected by the automated system but not by the informal system. Isolation counts for (a) *S*. Virchow and (b) *S*. Typhimurium PT 44 during 1993–4 are shown (solid line). A more characteristic year, 1994–5, is shown for comparison (dashed line).

https://doi.org/10.1017/S0950268898001939 Published online by Cambridge University Press

Table 3. *Positive predictive value for SPOT v2.0*

	Rows		Positive predictive value (%)
	Retrieved	Relevant	
1993	210	133	63
1994	260	137	53
1995	285	165	58

period, with the majority of these occurring in two adjacent suburbs over a 4-week period (Fig. 1*b*). The subsequent, more typical, year's data are shown for comparison (Fig. 1, dotted lines). Both these *Salmonella* species are common in the regions shown, and exhibit different characteristic distribution patterns: *S.* Virchow exhibits a characteristic summer increase each year, while isolations of *S.* Typhimurium PT 44 are typically spread more evenly through the year.

Positive predictive value

'Positive predictive value' is a measure of how well the system filters out extraneous material from the relevant events, and indicates the probable usefulness of a given surveillance report. The positive predictive values of the reports obtained using SPOT v2.0 in this study are shown in Table 3. As with sensitivity, the relevance standard is the retrospective evaluation of an epidemiologist.

As seen in Table 3, the positive predictive value of these reports is in the range of 53–63%. Most of the extraneous rows (false positives) are isolated rows, i.e. single weeks of elevated incidence with no elevated incidence in the weeks immediately preceding or following. As seen in Table 1, isolated weeks are easily differentiated from clusters by inspection.

Because the majority of false positives in a report are isolated rows, we hypothesized that the positive predictive value could be improved by excluding isolated rows. We experimented with adding a filter to exclude isolated rows from the SPOT report. This step improved the positive predictive value to 68, 74 and 68% for 1993, 1994 and 1995, respectively. Unfortunately, the filter also distorted the pattern of clusters. Some outbreaks are foreshadowed by an isolated week of increased incidence, and isolated weeks of high incidence may also occur at the tail end of an outbreak. In practice, the improvement in positive predictive value did not add significantly to the usefulness of the system, so the filter was not incorporated into SPOT v2.0.

In another experiment we set the heuristic cut-off used in calculating the warning threshold to 1. This resulted in increasing the sensitivity to almost 100%, relative to the epidemiologists' assessment, but the positive predictive value dropped to $< 20\%$. This level of extraneous material in the reports decreased their usefulness, so the heuristic cut-off was set to 2.

DISCUSSION

Surveillance is pivotal in the field of epidemiology, and early warning systems are a means whereby surveillance data can be translated into timely public health measures. We have developed an automated early warning system for routine surveillance and detection of outbreaks of infectious disease. As a first step, we have concentrated on gastroenteritis due to *Salmonella enterica*. The system design is general and can be extended to other organisms and other diseases where reasonably complete data are available.

There are certain requirements that an automated early warning system must satisfy if it is to be useful. In the case of salmonella, the model has to accommodate the large number of different strains of salmonella, each with its own characteristic pattern of geographic and seasonal occurrence. While individual modelling has been used with apparent success on a limited scale [15], this is not feasible for the > 700 different salmonella serovars in Australia and the large number of different distribution patterns. The seasonally fluctuating baseline rules out the use of the scan statistic [16], which has been used with success to detect clusters of trisomy but relies on a flat baseline [11, 17].

Another requirement for the model is robustness to aberrations in the data, since epidemiological databases invariably contain abnormal clusters in past data. In the case of food-borne illness, the abnormal clusters are often due to outbreaks in previous years, which may or may not have been recognized. Stroup and colleagues compare various statistical methods and note that a classical parametric method based on the mean of past data is not robust to aberrations [11]. The use of medians and a compound smoothing technique make our system more robust in the presence of outlying data.

Statistically-based methods can sometimes overcome the problem of robustness, but because the

https://doi.org/10.1017/S0950268898001939 Published online by Cambridge University Press

distribution of many salmonella serotypes is sparse over the geographic ranges of interest, many statistical methods are not appropriate to this application. A few methods have worked well for sparse data, notably the bootstrap method of stroup and colleagues [11], and the time series method of Watier and colleagues [15], but both are computationally intensive.

Computational efficiency is another requirement if a system is to be used regularly. Our combination of heuristics with a statistical method is computationally efficient. We have achieved additional improvement in real-time performance by separating the time-consuming baseline calculation from the surveillance, using a look-up table at surveillance time rather than calculating baselines on the spot. Consequently, real-time performance of our system takes only a few minutes and can therefore be done at any time. Because the actual surveillance is so fast, surveillance checks can be, and often are, run repeatedly as more data become available. Given that the reference date is the specimen date, repeat runs give increasingly accurate views of the pattern of incidence over the time period of interest, as more reports are received.

The real test of any surveillance system is how well it picks up outbreaks. We have previously reported informal observations of the effectiveness of an earlier version of the system, SPOT v1.0 [7] and made similar observations for the current version, SPOT v2.0. Striking examples of peaks that were first detected with the automated system are those of *S*. Virchow and *S*. Typhimurium, phage type 44, shown in Figure 1. Both these salmonella strains are relatively common in the respective regions, and the patterns of isolation suggest a single source outbreak on top of the normal background. Even these large excesses were not detected by the informal system over the noisy, fluctuating baseline. In the case of *S*. Virchow this might be considered surprising, since the peak is quite obvious in Figure 1. It must be remembered, however, that notifications arrive at different times after isolation and are processed by different staff. Under these conditions, isolations of serovars that are common in a geographic region do not necessarily raise the index of suspicion. Without a ready means of grouping and screening the data, even peaks that are obvious in retrospective analysis are easily missed at the critical time. As we had earlier hypothesized, the automated system outperformed the informal system in detecting abnormal clusters of common organisms.

While the informal evaluation gave us a sense that the automated system was useful, we sought a more formal way to quantitate the effectiveness of our system. We needed to choose a standard against which to evaluate our system. The goal of an early warning system is to warn of clusters in the available data that have a reasonable likelihood of indicating an outbreak. Because surveillance data are both incomplete and noisy, there may be outbreak associated data that are sparsely recorded in the database (e.g. an under-reported outbreak) and there may be suspicious looking clusters that are not really outbreaks (e.g. coincidence). An automated system is only as good as its input data. Therefore, rather than attempt to measure the system against 'true outbreaks', which are never known with any reliability anyway, we measured our system against epidemiologists' independent views of what their ideal warning system would flag as 'potential outbreaks' from the existing data. Because this standard relies on human judgement, the measurements are not absolute. Nonetheless, the epidemiologists' expert opinion provides a reasonable approximation to what is wanted. This kind of standard has been used with success in the discipline of information retrieval, where human judgement has been used as the standard for relevance of documents retrieved in response to database queries [18].

We have measured the sensitivity of our system and the positive predictive value of reports. These measures are recommended by the CDC for use in evaluating surveillance systems [14]. They are also widely used in the discipline of information retrieval, where the term 'recall' is equivalent to sensitivity and the term 'precision' is equivalent to positive predictive value. We know from the information retrieval literature that it is not generally possible to maximize both sensitivity (recall) and positive predictive value (precision) at the same time [18]. Each application must find the appropriate balance between these measures for its domain. Because an early warning system deals with human disease, we have set a higher priority on sensitivity than on positive predictive value, i.e. we were prepared to tolerate some level of extraneous data in our warning report, provided that we picked up most of the potential problems. These priorities influenced our choice of method and parameters.

With the method and parameters we describe here, we have achieved consistently $> 90\,\%$ sensitivity, with $\sim 50\,\%$ positive predictive value over 3 years. It is difficult to compare our system with the few other

systems available [8–11], because quantitative measures of effectiveness are often not reported. A system developed at the Communicable Disease Surveillance Centre, UK, had positive predictive value estimated at 40% [8]. In a recently reported system based on the quality control measure cumulative sum statistic, detection of salmonella outbreaks ranged between 0 and 100% for different states, with an overall specificity of 76–82% [9]. Our system compares favourably with these.

In practice, we have found $>90\%$ sensitivity, with $\sim 50\%$ positive predictive value a useful working range. The lower value for positive predictive value means that several rows in a SPOT report do not represent potential outbreaks. This does not actually translate into excessive unwarranted investigations, however, since the majority of the extraneous rows are isolated weeks, immediately obvious to the viewer. From a retrospective point of view, the sensitivity is close to the maximum of what can be expected from the available data. Eight clusters, or 'potential outbreaks', were not detected by SPOT in our study over 3 years, out of a total of 154 'potential outbreaks' nominated by the epidemiologists as worthy of investigation during this period. Of the 8 not detected, 6 were self-limited outbreaks of such small magnitude and such short duration that their early detection would have had no impact from the public health point of view, and the remaining two were due to aberrations in data entry. Real-time sensitivity would probably be somewhat lower than reported here, due to delays in processing specimens.

By using quantitative measurement, we have been able to confirm our previously reported informal observation that the introduction of the automated early warning system improved the rate of detection of outbreaks. While our measurements are not absolute, they give us a reasonable idea of how well our system performs. We suggest that more widespread use of standard measurements for surveillance systems will provide a useful means for comparing the effectiveness of different systems on different databases, and will do much to inform the choice of methods used in future applications. With more information available, these systems can continue to improve, and to deliver some of their potential in the area of public health.

ACKNOWLEDGEMENTS

We wish to thank Ms J. Powling of the Microbiological Diagnostic Unit for offering her expert views on the data. The statistical methods were integrated into the system by Ms Josipa Mickova, who was funded by a studentship from the Department of Computer Science of The University of Melbourne and by the Microbiological Diagnostic Unit of The University of Melbourne. NEPSS was funded by the Australian Health Ministers' Advisory Council.

REFERENCES

1. Fisher IST, Rowe B, Barlett CLR, GIll ON. 'Salm-Net' – laboratory-based surveillance of human salmonella infections in Europe. PHLS Microbiol Dig 1994; **11**: 181–2.
2. Salmon RL, Bartlett CLR. European surveillance systems. Rev Med Microbiol 1995; **6**: 267–76.
3. Baird-Parker AC. Foods and microbiological risks. Microbiology 1994; **140**: 687–95.
4. Hedberg CW, MacDonald KL, Osterholm MT. Changing epidemiology of food-borne disease: a Minnesota perspective. Clin Infect Dis 1994; **18**: 671–80.
5. Rodrigue DC, Tauxe RV, Row B. International increase in *Salmonella enteritidis*: a new pandemic? Epidemiol Infect 1991; **105**: 21–7.
6. Cohen ML, Fontaine RE, Pollard RA, Vonallen SD, Vernon TM, Gangarosa EJ. An assessment of patient-related economic costs in an outbreak of salmonellosis. N Engl J Med 1978; **299**: 459–60.
7. Stern L, Lightfoot D, Forsyth JRL. Automated detection of salmonella outbreaks. In: Witten M, ed. Proceedings of the First World Congress on Computational Medicine, Public Health, and Biotechnology. Singapore: World Scientific, 1996: Part III, 1395–404.
8. Farrington CP, Andrews NJ, Beale AD, Catchpole MA. A statistical algorithm for the early detection of outbreaks of infectious disease. J R Stat Soc, Series A 1996; **159**: 547–63.
9. Hutwagner LC, Maloney EK, Bean NH, Slutsker L, Martin SM. Using laboratory-based surveillance data for prevention: an algorithm for detecting *Salmonella* outbreaks. Emerg Infect Dis 1997; **3**: 395–400.
10. Stephenson J. New approaches for detecting and curtailing foodborne microbial infections. JAMA 1997; **277**: 1337–40.
11. Stroup DF, Williamson GD, Herndon JL. Detection of aberrations in the occurrence of notifiable diseases surveillance data. Stat Med 1989; **8**: 323–9.
12. Velleman PF, Hoaglin DC. Applications, basics, and computing of exploratory data analysis. Boston: Duxbury Press, 1981.
13. Minitab Reference Manual. Valley Forge, PA: DataTech Industries, 1989.
14. Klaucke DN, Buehler JW, Thacker SB, Gibson RG,

Trowbridge FL, Berkelman RL. Guidelines for evaluating surveillance systems. MMWR 1988: **37**(S-5): 1–17.
15. Watier L, Richardson S, Hubers B. A time series construction of an alert threshold with application to *S. bovismorbificans* in France. Stat Med 1992; **10**: 1493–509.
16. Naus JL. The distribution of the size of the maximum cluster of points on a line. J Am Stat Assoc 1965; **60**: 532–53.
17. Wallenstein S. A test for detection of clustering over time. Am J Epidemiol 1980; **111**: 367–72.
18. Korfhage RR. Information storage and retrieval. New York: John Wiley and Sons, 1997.

https://doi.org/10.1017/S0950268898001939 Published online by Cambridge University Press