JOURNAL OF CLINICAL MICROBIOLOGY, Apr. 1993, p. 857–860
0095-1137/93/040857-04$02.00/0

Vol. 31, No. 4

Computerized Surveillance in Clinical Microbiology with Time Series Analysis

RAM BENNY DESSAU* AND PETER STEENBERG

Department of Clinical Microbiology, Herlev University Hospital, Herlev Ringvej, Denmark 2730

Received 7 October 1992/Accepted 11 January 1993

An automatic surveillance system to detect changes in the incidences of microorganisms diagnosed in the department of clinical microbiology has been developed. The program is incorporated into the laboratory computer system and gives a weekly list of microorganisms whose isolation rates compared with those of a previous period exceed a chosen limit. The system uses time series analysis with moving weighted averages, and the detection limit is based on the distribution of the residuals. Output from the system included information about potential outbreaks of gastroenteritis, nosocomial infection with *Corynebacterium jeikeium*, and a seasonal epidemic of respiratory syncytial virus. The system also listed organisms not commonly isolated in the laboratory and detected incorrect reports. We conclude that continuous surveillance of laboratory data with time series analysis is a valuable tool for epidemiologic surveillance and quality control. Large quantities of data may be screened.

In 1991, the department of clinical microbiology in Herlev County Hospital received 196,000 specimens for testing. It is therefore not possible for the medical staff to gain an overview of activity by manually reviewing reports except for those on blood cultures and other specimens of special importance.

We have implemented a time series analysis model to screen large amounts of data continuously for changes in the incidences of specific microorganisms reported from the department. Observations of a phenomenon which is moving through time generate an ordered set known as a time series (3). Time series analysis (3, 6) involves a wide range of statistical techniques used extensively in agriculture, industry, economics, and engineering for analysis of previous developments and for forecasting.

MATERIALS AND METHODS

Computer registration. All specimens are routinely registered in the microbiological data base system (ADBakt, AUTONIK AB; RAMSTA, Sköldinge, Sweden) running on a VAX 4200 minicomputer with 40 terminals connected. The data base system is programmed in MUMPS and is presently used in 13 Swedish laboratories and 2 Danish laboratories. The system consists of a large number of procedures covering almost all aspects of microbiology from registration of the specimen for testing to the final report to the physician. A wide range of listings, summary reports, and statistics is available.

Data collection. The surveillance system described below was programmed in MUMPS by us and incorporated into the ADBakt system.

Microbiological diagnoses from final reports are counted weekly and stored in a separate data file as (i) total number of positive specimens with a specific microbiological name and (ii) number of patients with a specific microbiological diagnosis. Specimens from one patient that contain the same organism are counted only once within 3 weeks. Furthermore, the data are subdivided according to the customer (hospital department, private clinic, etc.) and type of specimen (blood culture, urine culture, etc.). These options have not been used in the current application. Other laboratory data such as resistance patterns are not used in the present setting. Each year, the weeks are numbered from 1 to 52 starting either with the Monday at the end of December of the preceding year or with January 1 depending on the calendar.

The principle is to choose a basic period and count the number of events within each period. Then choose the number of periods (n) and the weights ($w_1 \ldots w_n$). Given the observed count of events up to time ($u_{a-n+1} \ldots u_a$), where a is the current period, we can calculate the weighted average (w_a) and the residual (r_a) as follows:

$$w_a = \left(\sum_{i=1}^{n} w_i\right)^{-1} \sum_{i=1}^{n} w_i u_{a-n+i}$$

$$r_a = u_a - w_a$$

Thus, the residual is the weighted average (the expected count) subtracted from the observed count. This procedure is repeated for n_r consecutive periods; hence the term "moving weighted averages." The standard deviation (s_a) of the residuals, r_a, and the standardized normal deviate (z_a) are calculated (1) as follows:

$$s_a = \sqrt{\frac{\Sigma(r_i - \bar{r})^2}{n_r - 1}}$$

and

$$z_a = \frac{r_a}{s_a}$$

Upper and lower detection limits are chosen on an empiric basis as a value of z. If the normal deviate (z_a) exceeds the detection limit, the microorganism is included on the surveillance list.

* Corresponding author.

TABLE 1. Calculation of moving weighted averages for *H. influenzae* from week 19 (May) to week 37 (September), 1991[a]

Week	u_a	w_a	r_a	s_a	z_a
19	44				
20	44				
21	72				
22	60				
23	36				
24	39				
25	53	52.5	0.5		
26	44	48.4	−12.4		
27	38	45.3	−7.3		
28	34	42.2	−8.2		
29	21	39.7	−18.7		
30	28	36.8	−8.8		
31	25	32.2	−7.2		
32	35	29.2	5.8		
33	15	27.1	−12.1		
34	24	26.8	−2.8		
35	29	25.7	3.3		
36	22	24.7	−2.7		
37	48	25.5	22.5	10.4	2.2[b]
38	35	26.7	8.3	10.8	0.8
39	53	31.6	21.4	12.4	1.7

[a] u_a, weekly number of isolates; w_a, weighted average; r_a, residual; s_a, standard deviation; z_a, normal deviate.
[b] 2.2 > 1.96 (detection limit).

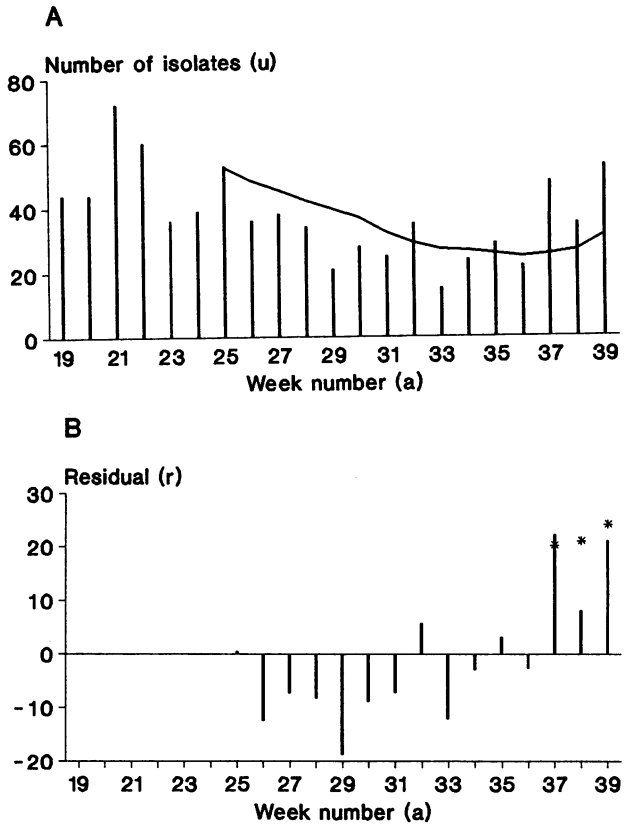

FIG. 1. Data from Table 1. (A) Weekly counts of isolates containing *H. influenzae* isolated from any specimen for weeks 19 to 39 in 1991 (vertical bars) and weighted moving averages (w) (horizontal curve). (B) Corresponding residuals (vertical bars). The residual in week 37 is just exceeding the detection limit (*).

RESULTS

Choosing parameters for the calculation. The basic period of 1 week was chosen because the number of final reports show a weekly periodicity, with a maximum from Tuesday to Friday and a minimum during the weekend. Seven weeks (n) and the weights 1, 3, 5, 6, 5, 2, and 1 were chosen to calculate the moving average (w_a). Thirteen residuals (n_r) were used to calculate the standard deviation (s_a) of the residuals, and the detection limit was set at $z = 1.96$. Simulations on retrospective data showed that these choices were a suitable compromise to filter out random variation, slow seasonal variation, and long-term trend.

To illustrate the calculation, the weekly number of isolates with *Haemophilus influenzae* is shown (Table 1 and Fig. 1). Data for 19 weeks are needed to derive the first standard deviation.

Output. The weekly list (Table 2) shows a number of rare microbiological diagnoses and a few more-common organisms. The "normal deviate" gives an impression of the magnitude of the deviation. Uncommon organisms which have not occurred previously during the period have a deviate of 3.60. After reading the lists and comparing the number of specimens with the number of patients, one can choose an item to investigate further.

Epidemiological surveillance. The following results are chosen to illustrate the behavior of the surveillance system in a wide range of situations over a period of 1 year.

(i) Large number of events. Pneumococci (Fig. 2) are common organisms in respiratory tract secretions at a fairly constant level all year round. This is an example of simulation with retrospective data. The detection limits are continuously adjusting to the current variation.

(ii) Large seasonal variation. Respiratory syncytial virus (Fig. 3) is common in children during winter from November to March. A few cases were diagnosed and exceeded the limit in week 44, in October 1991. They were brought to our attention again in week 4, in January 1992, at the beginning of the epidemic.

(iii) Low number of events. Meningococci (Fig. 4) were isolated in low numbers during 1990. The diagnoses of two and three patients with meningococci during two consecutive weeks around New Year's Day 1991 was brought to our attention by the system. In this instance, the cases did not constitute an epidemic of meningitis. Two of the isolates were from routine throat cultures, and there was no connection between the three cases of meningitis.

In Table 2, an increasing number of patients with *Salmonella enteriditis* is indicated. On further investigation, we found a greater increase than during the same period last year in our department. In this case, we decided to monitor the development.

Some other events of interest detected over a full year from September 1991 included a small outbreak of *Salmonella typhi* within the same family that was automatically detected by the system. Three cases of infection with multiresistant *Corynebacterium jeikeium* were detected in the department of haematology, where the clinicians were already aware of two of the cases but not of the third one. An investigation was performed by the hospital infection team, and all three patients were found to have been on the same ward. The system detected an increase (deviate = 2.78) in the number of patients with group G streptococci. The follow-up investigation showed a simultaneous increase in

TABLE 2. Listing for Week 28 in July 1992[a]

Microorganism	No. of patients		
	This week	Last week	Normal deviate
Streptococcus lactis	1	0	3.60
Pseudomonas alcaligenes	1	0	3.60
Escherichia hermanii	1	0	3.60
Salmonella typhimurium	4	2	2.55
Salmonella infantis	2	0	3.60
Salmonella enteriditis	12	5	2.91
Aeromonas hydrophila	1	0	2.31
Pasteurella species	1	0	3.24
Bacteroides fragilis group	13	2	2.46
Bacteroides fragilis	2	0	2.09
Candida guilliermondii	1	0	3.60
Aspergillus niger	1	0	3.24
Chilomastix mesnili	1	0	3.60
Plasmodium malariae	1	0	3.15
Pneumocystis carinii	1	0	2.10
Motile gram-negative rods (*Pseudomonas* spp.?)	2	1	2.36
Gram-positive cocci, 10,000/ml of urine	12	4	2.19
Mixed flora, 1,000/ml of urine	19	6	2.18

[a] Date of listing, July 15, 1992; the period used in the calculation of the normal deviate is the week of October 28, 1992. Length of retrospective period, 7 weeks; total length of period, 7 weeks; detection limit, 1.96; period for calculation of standard deviation, 13 weeks; weights: 1, 3, 5, 6, 5, 2, and 1; sum, 23.

isolation of this organism from different hospital departments and types of specimens with no apparent epidemiological connection. The matter was regarded as stochastic variation.

Quality control. The system lists microbiological diagnoses that are uncommon in our laboratory. In the laboratory most types of bacteria are indentified with the API system (Biomérieux SA). This system is able to identify a large number of bacteria with a variable probability of correct identification. In a few cases, peculiar microorganisms included on the weekly list led to reexamination of the isolates. In week 40, 1991, the system detected incorrect reports for beta-hemolytic streptococci without serotyping according to Lancefield groups. This resulted in a modification of the ADBakt system so that the laboratory staff could not report a beta-hemolytic streptococcus without grouping for the final report.

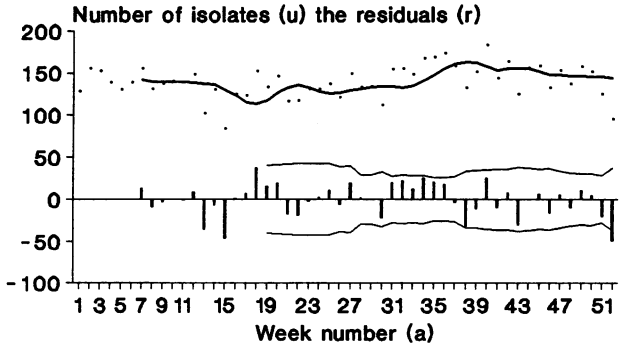

FIG. 2. Weekly numbers of isolates containing *Streptococcus pneumoniae* isolated from any specimen in 1990, weighted moving average (——) (*w*), and (below) residuals (vertical bars) with upper and lower detection limits (——) (*z*). The very few isolates in the last week of the year exceeded the detection limit, because fewer specimens are submitted during the holidays.

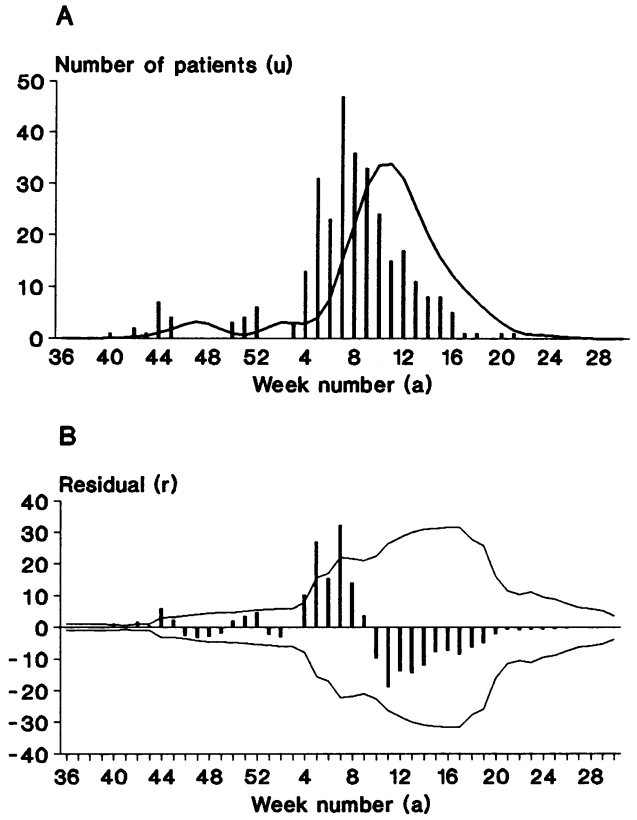

FIG. 3. (A) Weekly numbers of patients (vertical bars) with respiratory syncytial virus from August 1991 (week 36) to July 1992 (week 30), with moving averages (——) (*w*). (B) Residuals (vertical bars), with upper and lower detection limits (——) (*z*).

DISCUSSION

Surveillance of changes in positive culture rates for identifying potential outbreaks of nosocomial infections has been done by others (2, 5) by using a Poisson distribution. Our data did not fit the Poisson distribution as determined by the Poisson heterogeneity test (1) because of seasonal long-term trends, changes in random variation and mean values as a function of time, and the type of microorganisms surveyed. In another model (7), excessive culture rates were defined as greater than or equal to twice the monthly baseline rate, which was derived from the mean of positive culture results for the previous 12 months. A positive culture rate greater than or equal to four times the mean rate was associated with significantly greater specificity for identifying problems of nosocomial infection. This model does not adjust to seasonal variation, and the system described here is thus more sensitive and could detect relevant information from about two times the mean rate of the preceding period. Two outbreaks (4) of nosocomial infection with *pseudomonas* and *proteus* spp. were detected by choosing culture rates exceeding a fixed threshold of 80% based on the number of isolates found during a single 26-week period. This choice generates a stiff threshold line which is not continuously

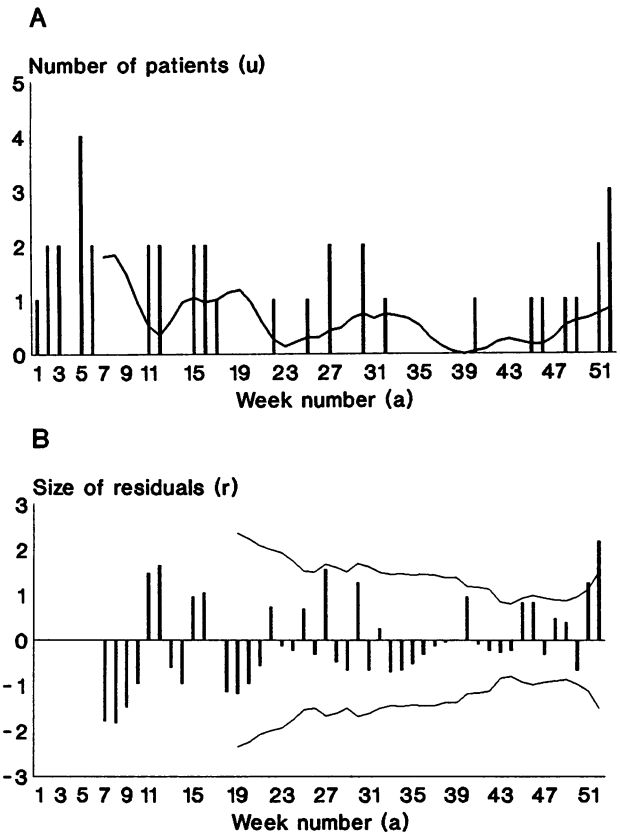

FIG. 4. (A) Weekly numbers of patients (vertical bars) with *Neisseria meningiditis* in 1990. —, moving average (*w*). (B) Residuals (vertical bars), with upper and lower limits. ——, detection limit (*z*).

updated by a moving principle and thus does not adapt to seasonal or long-term change.

The distribution of residuals for the smaller counts is skewed (Fig. 4). As the number of events approaches zero, deviations above the moving average become greater than those below. Therefore, the standard deviation is merely a manner of calculation and a numerical expression of the relative size of the residual compared with our chosen limit and may not validly be translated to *P* values in a statistical table. The limit of $z = 1.96$ was chosen out of habit, but it gave a practical amount of information for between 8 and 25 events on the weekly list and a reasonable sensitivity as judged by retrospective data. Future experience will undoubtedly show the need for adjustment. The use of symmetric detection limits on the asymmetric distribution of residuals was an advantage, as we were not interested in a decrease unless it was very large.

Specificity may be increased by choosing more-specific items to be monitored. In nosocomial infection surveillance, one could choose to monitor certain departments, e.g., the intensive care unit or the department of hematology. Surveillance of certain antimicrobial resistance patterns is another area of interest which could be implemented by using the same model. Adjusting the weights and the length of the retrospective period determines the fit of the moving average to the observed count. We wanted to monitor unexpected short-term changes and therefore used a period of 7 weeks. Changing these parameters might adapt the model to different data sets. The weights may be symmetrical or asymmetrical, and negative values may also be used. Longer periods can be chosen for monitoring long-term variation by taking the sum of any number of weeks, e.g., one-quarter of a year, or approximately 13 weeks.

We have used a retrospective period of 13 weeks for calculating the standard deviation of the residuals, thus making the chosen limit slower to respond to changes in variance. When there is extreme seasonal variation, the large distance (Fig. 3B) between the detection limit curves goes well into the summer season before adjusting close to the zero line. In this case, a smaller postepidemic accumulation would not have been detected, and a shorter period for calculation of the standard deviation could have been more appropriate.

Thus, the described system of continuous automatic surveillance can screen large quantities of data for changes in the incidence of microorganisms detected in the clinical microbiology laboratory. Sensitivity, specificity, and periodicity may be adjusted to local conditions and the specific task.

ACKNOWLEDGMENT

We were kindly assisted in this work by the Department of Data Processing, Herlev University Hospital.

REFERENCES

1. **Armitage, P., and G. Berry.** 1987. Statistical methods in medical research, 2nd ed. Blackwell Scientific Publications, Oxford.
2. **Hansen, L., H. J. J. Kolmos, and K. Siboni.** 1978. Detection of cumulations of infections in hospital over a three-year period using electronic data processing. Dan. Med. Bull. **25**:253–257.
3. **Kendall, M. G., and A. Stuart.** 1966. The advanced theory of statistics, vol. 3. Charles Griffin & Co. Ltd., London.
4. **McGuckin, M. B., and E. Abrylun.** 1979. A surveillance method for early detection of nosocomial outbreaks. A.P.I.C. J. Infect. Control **7**:18–21.
5. **Møller, J. K., P. Bülow, O. J. Bergmann, and J. Ellegaard.** 1989. Accumulated microbiological data. Surveillance of infections/antibiotic policy. Ugeskr. Laeg. **151**:1834–1837.
6. **Montgomery, D. C., L. A. Johnson, and J. S. Gardiner.** 1990. Forecasting and time series analysis, 2nd ed. McGraw-Hill Book Co., New York.
7. **Shifman, R. B., and R. A. Palmer.** 1985. Surveillance of nosocomial infections by computer analysis of positive culture rates. J. Clin. Microbiol. **21**:493–495.