Eval of Fig 1 .pdf

American Journal of Epidemiology

Vol. 137, No. 3
Printed in U.S.A.

Evaluation of a Method for Detecting Aberrations in Public Health Surveillance Data

Donna F. Stroup,[1] Melinda Wharton,[2] Karen Kafadar,[3] and Andrew G. Dean[1]

The detection of unusual patterns in routine public health surveillance data on diseases and injuries presents an important challenge to health workers interested in early identification of epidemics or clues to important risk factors. Each week, state health departments report the numbers of cases of about 50 notifiable diseases to the Centers for Disease Control and Prevention, and these reports are published weekly in the *Morbidity and Mortality Weekly Report*. A new analytic method and a horizontal bar graph were introduced in July 1989 to facilitate easy identification of unusual numbers of reported cases. Evaluation of the statistical properties of this method indicates that the results are fairly robust to nonnormality and serial correlation of the data. An epidemiologic evaluation of the method after the first 6 months showed that it is useful for detection of specific types of aberrations in public health surveillance. *Am J Epidemiol* 1993;137:373–80.

epidemiologic methods; public health surveillance

A foundation of the science of epidemiology is the study of the departure of observed health event experience from the expected occurrence (1). The detection of unusual patterns in reports of diseases and injuries presents an important challenge in public health surveillance. Aberrations in usual distributions of health event occurrence in different geographic areas or different time periods may provide an early signal of an epidemic or important clues to the etiology of the disease or to specific risk factors for the event (2).

A critical concept in epidemiology is the comparison of an observed number of disease or injury reports with what is usual or normal. We use the term *aberration* here to denote changes in the occurrence of a health event that are statistically significant when compared with usual or normal history. The existence of an aberration is necessary, but not sufficient, for the occurrence of an *epidemic*; thus, false-positive findings, as well as false-negative results, are a concern. For example, the Centers for Disease Control and Prevention defines an excess in influenza-related mortality only when the number of reported deaths due to pneumonia and influenza exceeds a 95 percent confidence limit in the forecast for two or more consecutive periods (3). Definition of an influenza epidemic, however, requires laboratory data and further epidemiologic evidence of influenza-related morbidity. Indeed, an influenza epidemic can occur in the absence of excess mortality. Statistical methods are intended for routine use by the public health analyst in conjunction with epidemiologic investigation and close communication with the source of the surveillance reports.

State health departments report the numbers of cases of about 50 notifiable diseases

Received for publication February 3, 1992, and in final form June 29, 1992.

[1] Division of Surveillance and Epidemiology, Epidemiology Program Office, Centers for Disease Control and Prevention, Atlanta, GA.

[2] National Center for Prevention Services, Centers for Disease Control and Prevention, Atlanta, GA.

[3] Division of Cancer Prevention and Control, National Cancer Institute, Bethesda, MD.

Reprint requests to Dr. Donna F. Stroup, Division of Surveillance and Epidemiology, Epidemiology Program Office, Mail Stop C08, Centers for Disease Control and Prevention, 1600 Clifton Road NE, Atlanta, GA 30333.

each week to the National Notifiable Diseases Surveillance System of the Centers for Disease Control and Prevention. The list of health events is determined collaboratively by the Council of State and Territorial Epidemiologists and the Centers for Disease Control (4, 5).

Each week, provisional reports are published in the *Morbidity and Mortality Weekly Report* and are made available to epidemiologists, clinicians, and other public health professionals in a timely manner. Although the tables of the *Morbidity and Mortality Weekly Report* provide important information, the volume of data and the need for ease of interpretation encourage the development of a graphic display to highlight unusually high or low numbers of reported cases.

A new analytic and graphic method was developed to achieve the following objectives: 1) to depict in a single comprehensible graph weekly reports of approximately 20 disease totals that can be compared with past results; and 2) to highlight for further analysis the results most likely to indicate changes in long-term trends or epidemics. These objectives were formulated to reflect most recent behavior in as short a time period as possible for weekly publication, but long enough to provide stable results. To facilitate comprehension, we used the same method for all diseases represented graphically.

This paper presents an evaluation of the method as a tool for the routine analysis of public health surveillance data. We summarize a statistical evaluation of the method, as well as an epidemiologic assessment, for national surveillance data during the first 6 months of its use. We also consider additional issues that arise when the method is used for state data and discuss alternative methods that may be used for the detection of aberrations.

MATERIALS AND METHODS

The analytic method currently used as figure I in the *Morbidity and Mortality Weekly Report* (figure 1), called the Morbidity and Mortality Weekly Report Current/Past Experience Graph of the Centers for

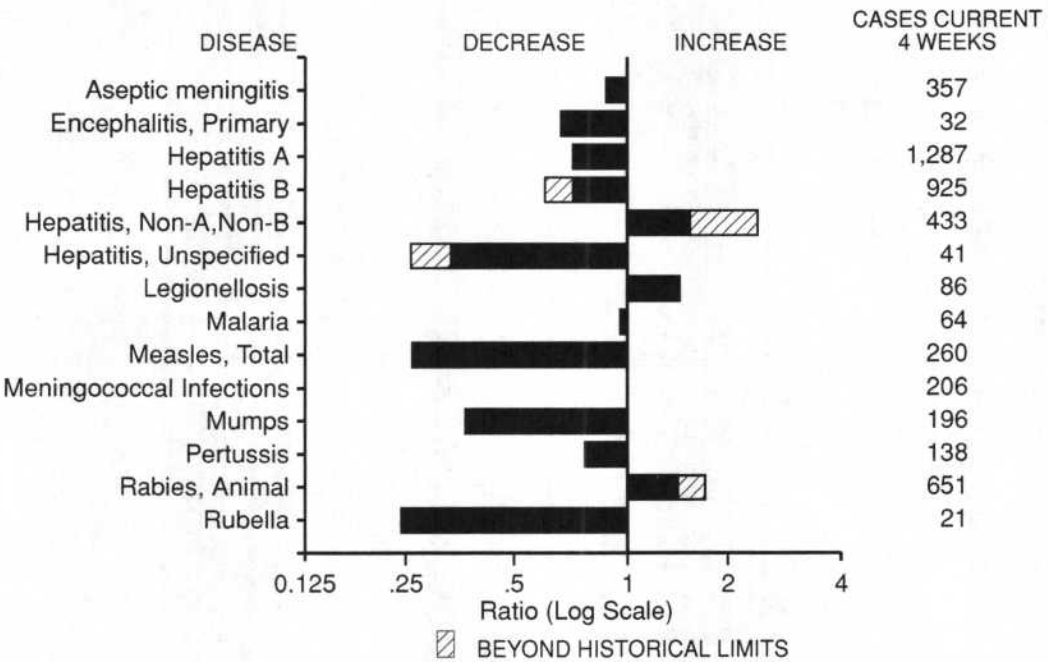

FIGURE 1. Notifiable disease reports, comparison of 4-week totals ending May 23, 1992, with historical data, United States. The ratio is of current 4-week total to the mean of 15 4-week totals (from previous, comparable, and subsequent 4-week periods for the past 5 years). The point where the hatched area begins is based on the mean and two standard deviations of these 4-week totals.

Disease Control and Prevention, compares the number of reported cases in the current 4-week period for a given health event with historical data from the preceding 5 years (6, 7). Numbers of cases in the current 4-week period are listed to facilitate interpretation of instability caused by small numbers.

The choice of 4 weeks as the "current period" was based on evidence of weekly fluctuation in disease reporting that is usually due to irregular reporting rather than to disease incidence. The use of a 5-year history achieves the objective of applying the same model for all conditions depicted. This is particularly helpful since some health events were made notifiable only recently (e.g., acquired immunodeficiency syndrome and legionellosis). In addition, modeling of reported influenza incidence has shown that more accurate forecasts are based on more recent data (8). To increase the historical sample size and to account for any seasonal effect, the baseline is taken to be the average of the reported number of cases for the preceding 4-week period, the corresponding 4-week period, and the following 4-week period, for the previous 5 years. This yields 15 correlated observations, referred to as the historical observations or baseline (figure 2).

The deviation from unity of the ratio of the current total to the historical average is indicative of a departure from past patterns.

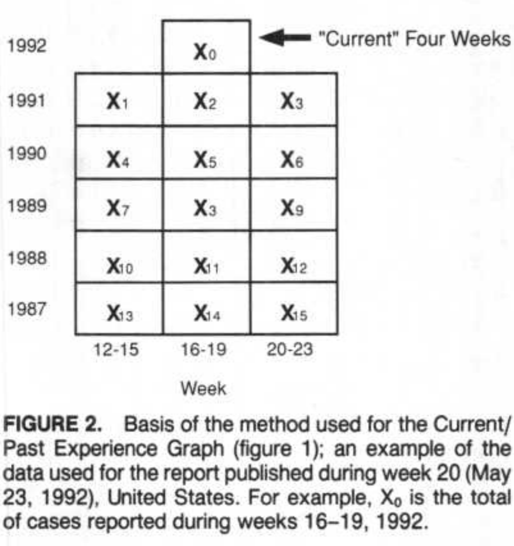

FIGURE 2. Basis of the method used for the Current/Past Experience Graph (figure 1); an example of the data used for the report published during week 20 (May 23, 1992), United States. For example, X_0 is the total of cases reported during weeks 16–19, 1992.

We plot this ratio on a logarithmic scale so that an *n*-fold increase projects to the right the same distance as an *n*-fold decrease projects to the left, and no change from past patterns (1:1) produces a bar of zero length (9). To distinguish the conditions that may require further investigation, the hatching on the bars begins at a point based on the mean and standard deviation of the historical observations (6). (Historical limits of the ratio of current reports to the historical mean are calculated as 1 ± 2 times the standard deviation divided by the mean, where the mean and the standard deviation are calculated form the 15 historical 4-week periods.)

RESULTS

Statistical evaluation

Because surveillance data are reported sequentially in time, they may not satisfy the assumptions necessary for usual time series analyses. For example, the number of measles cases reported to the Centers for Disease Control and Prevention in a given 4-week period over the 5 years, 1985–1989, is highly correlated from period to period (Pearson product-moment correlation = 0.86). The problem is particularly apparent with incidence data for which the numbers of reported cases are subject to seasonal effects and reporting delays. The method used to set the hatching point for the Current/Past Experience Graph may be affected by the correlation of reported health events over time, and little is known about the empirical performance of the method in the presence of such correlation.

To investigate this problem, we performed a simulation study (10) in which the values needed for the ratio comparison were simulated from a known distribution. The purpose of this simulation was to provide a "gold standard" or true confidence limit to compare with the limit produced by the method of the Current/Past Experience Graph. In addition to this method, we investigated three alternative methods for estimating the standard error of this ratio: the bootstrap (6), the jackknife, and the delta methods. The bootstrap method has been developed for the case of independent and identically distributed variables (11) and extended to the case when observations form a first-order autoregressive process (12) or when observations are independent but not identically distributed (13). The jackknife is preferred in situations in which groups of observations may be expected to be reasonably independent, even if individual observations are not (14), a situation that may seem appropriate here since the method of the Current/Past Experience Graph is based on groups of reports, each of which covers 4 weeks. Künsch (15) applies this concept of grouping to bootstrap calculations. The delta method is based on a first-order Taylor series expansion of the statistic; the variance of the ratio is computed from estimates of first and second moments of the statistic.

To evaluate these methods for providing the variance needed for the ratio of the Current/Past Experience Graph, we used a model by which results can be compared with true or model-simulated values. In addition, we applied the three methods to data from the National Notifiable Diseases Surveillance System for measles incidence and measured performance by epidemiologic confirmation of increased activity. A complete description of the model and the methods is found in the study by Kafadar and Stroup (10).

Of the four methods investigated (bootstrap, jackknife, delta, and the Current/Past Experience Graph), the delta method produced the best estimate of the "true" confidence bands; that is, delta calculations had the mean value closest to the simulated target for all values of period-to-period correlation. The method of the Current/Past Experience Graph performs almost as well as does the delta method, with departures from the reference value rarely more than 0.90, a 10 percent error.

The simple bootstrap (based on random sampling from the 15 historical observations) and jackknife methods produce overly optimistic (low) estimates of the variance used for the ratio in the Current/Past Experience Graph; the difference is at least an order of magnitude and sometimes more (e.g., when adjacent periods have a correlation of 0.1, the variance is only 4 percent of the simulated reference value). Such an underestimate of variance will result in a confidence band (hatching point) too close to one; thus, excesses for disease will appear too often, producing an overly sensitive system.

The four estimators were applied to the reported incidence of measles from 1980 to the first quarter of 1990. Most methods consistently identify most of 1986 and 1989 as unusually high for measles incidence; the delta method also identifies the last two periods of 1989 and the second period of 1990 as elevated (figure 3).

Epidemiologic evaluation

During the first 6 months (April 1, 1990 to September 29, 1990) of the implementation of the Current/Past Experience Graph, we received supplemental information from state health departments for conditions that were shown to be reported in excess using this method (16). For each "episode" (a reporting increase for 1 or more weeks), we were able to identify a few states that produced the excess. An exception was measles, which exceeded historical limits each week after April 21, 1990, with increases reported in most states. Usually, we were readily able to determine an explanation for the increase. These explanations included batch reporting and improved laboratory-based reporting as well as true outbreaks. (Batch reporting occurs when disease reports accumulated over a period of time are forwarded as one batch to the Centers for Disease Control and Prevention.)

For conditions with relatively small numbers of cases reported nationally (i.e., pertussis, legionellosis, and rubella), small increases in reporting resulted in excesses. Many of the reports from state health departments represented small outbreaks; yet, when aggregated with other nationally reported events, the excess was apparent. However, an outbreak of rubella highlighted

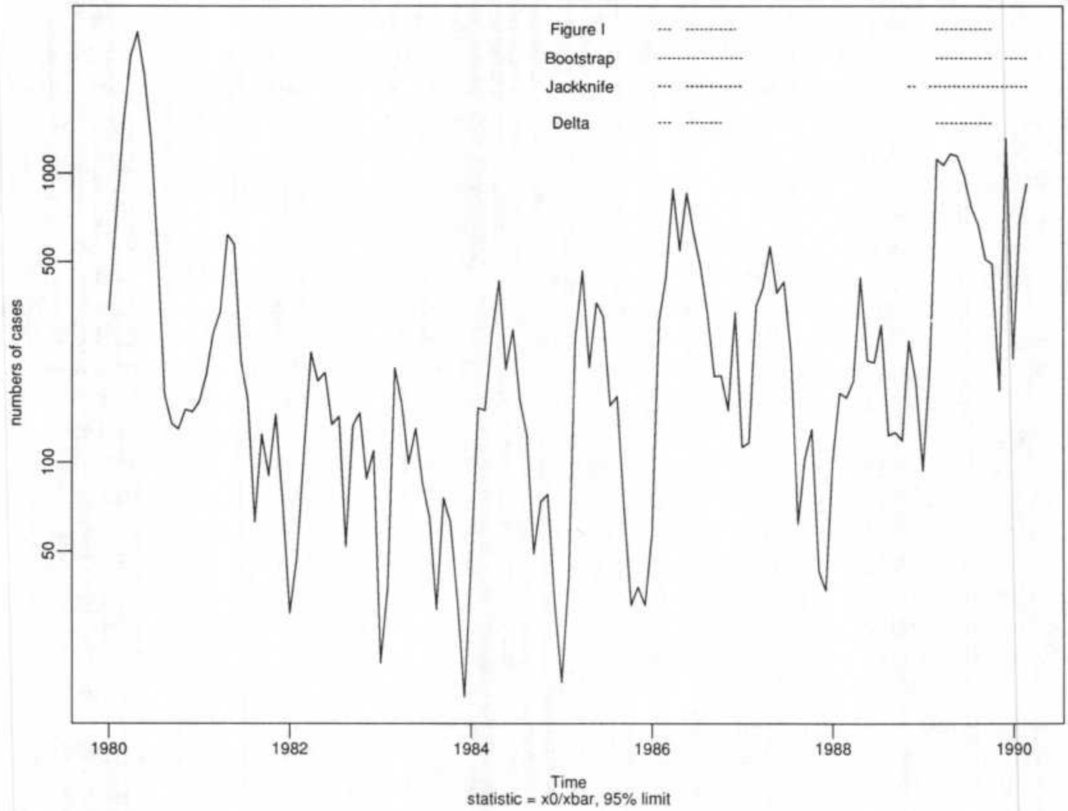

FIGURE 3. Reported cases of measles in the United States, 1980–1990 (first quarter). Source: Centers for Disease Control and Prevention, National Notifiable Diseases Surveillance System.

by this method proved to be of substantial public health importance (17).

DISCUSSION

No single method can be used to detect all epidemics or all types of aberrations. In spite of the known limitations of routine national surveillance data for notifiable diseases, such as incomplete or inaccurate reporting (18, 19), the data can be useful for demonstration of trends and for the detection of changes from historical patterns that signal the need for public health intervention. We discuss these issues for the method of the Current/Past Experience Graph.

What is the purpose of the surveillance system?

The data used for these analyses are reported weekly by state health departments. Although each state analyzes its own data, patterns apparent from the aggregated national picture may aid prevention and intervention efforts. Additionally, the data are maintained historically for the archival purposes of measuring trends and assessing the effects of interventions.

What is the purpose of the analytic method?

Since a single method cannot be expected to distinguish between a change in historical trend and a one-time outbreak with an unsustained level, the analyst must identify the purpose of the analysis before choosing an analytic method. If the nature of the data is determined and the questions are well defined, then the results of the analytic method can be used to augment other sources of information.

The purpose of the Current/Past Experience Graph is to facilitate the routine analysis of surveillance data and to supplement other sources of information. The method may not be useful for conditions with long-term historical trends, and for these we have proposed an alternative display (7).

When the data involve extensive patterns, such as the disease incidence data illustrated here, simplification of the patterns may be helpful. The classic methods of time series analysis are appropriate for this situation, but these may not be accessible to the practicing public health official. Often, very simple methods suffice. For reducing the magnitude of intrasample correlation on disease incidence data by 4-week period, the removal of period effects is apparently as convenient as the removal of long-term and seasonal trends by more complex time series methods such as those proposed in the statistical literature (20). Given knowledge of the correlation structure among successive groups of reports, estimates of the variance calculated from this model are more promising. However, the requirement that the correlation structure of each disease series be known requires continuous monitoring, which is impractical for most surveillance systems.

Which conditions should be monitored?

Any analytic method for routine surveillance is useful only for conditions associated with direct public health intervention. The method of the Current/Past Experience Graph is most appropriate for diseases that do not exhibit frequent changes in trend or level and that occur often enough so that a few cases do not constitute a significant flag. If the data are not preanalyzed for trend and period effects and the variance of the numerator (present cases) can be assumed to be the same as the variance of the observations in the denominator (historical data), the method of the Current/Past Experience Graph may be less powerful. This is especially true if the series exhibits considerable correlation for first-order (adjacent observations) processes and beyond. For rare conditions, the instability caused by small numbers of reported cases may make the results unsuitable for repeated use.

What is the unit of analysis?

We chose national data for presentation of the Current/Past Experience Graph. The objective was to use as short a time period as possible for weekly publication, thus making the results useful for timely intervention. However, variability in weekly reports caused by factors other than the disease process, such as reports delayed because of outbreaks, made the results unstable. Thus, we chose a 4-week window.

Because of the interest in analytic techniques for the analysis of aberrations in surveillance data at the state level, we undertook a study of the method of the Current/Past Experience Graph with six state health departments (21). During the 4-month period of study, a total of 210 episodes were observed, of which 27 were flagged as exceeding historical limits; one state had no episodes of unusual reporting. Overall, 14 episodes (52 percent) represented epidemiologically confirmed outbreaks. Many were small, and none were detected when aggregated with other state data for the national analyses. Each disease exceeded historical limits at least twice during the study period, and, for all but meningococcal disease, at least one of these represented an outbreak, as defined by the state health department. Although clearly the numbers are small, the proportion of episodes that represented outbreaks varies. This is expected for conditions with differing epidemiology.The five outbreaks known to the health department that were not detected by the method of the Current/Past Experience Graph highlight some of its limitations. In three outbreaks, cases were not reported to the Centers for Disease Control and Prevention as current reports; thus, they were not included with the data used for the calculation. The other two outbreaks were not detected because of increases in the corresponding baseline.

What provision is there for updating or correcting the data by using later reports?

In the National Notifiable Diseases Surveillance System, the Centers for Disease Control and Prevention encourages early report of a case and then allows for later confirmation or modification. The methodology of the Current/Past Experience Graph is applied to the provisional (earliest reported) data. In our study of six states, two of five outbreaks were not detected because of late reports not included in the current reporting period.

How is the baseline determined?

The choice of 5 years as a baseline period was based on a consideration of appropriate sample size balanced by a desire to use the same method for all conditions. Although a longer baseline may be possible for some conditions with a long reporting history, epidemics or changes in trend will increase the variance of the baseline and thus offset any benefit of additional data. In this case, the analyst may choose to omit or adjust the epidemic data.

How are outbreaks in the baseline handled?

The Current/Past Experience Graph as presented here does not adjust for epidemics which occur during the baseline period. The result of this is a progressive decline in sensitivity as the outbreak moves in and then out of the baseline window, as for measles. To address this point, one could use a median of the baseline reports (rather than a mean), such as we presented in the statistical evaluation section of this paper. Unfortunately, this replacement invalidates the technique used to compute the point for signaling aberrations, and the alternative methods for calculating this are not as accessible as the method of the Current/Past Experience Graph to the practicing epidemiologist.

What is the sensitivity and predictive value positive of the method?

This analysis shows that the sensitivity of the Current/Past Experience Graph is quite high. In part because of the use of provisional data, we use the mean in the calculation. We investigated positive predictive value by applying the Current/Past Experience Graph to data from six state health departments and asking each to follow up on aberrations detected by this system. In addition, we asked that outbreaks that came to their attention through other sources, but that were not identified by the Current/Past Experience Graph, be noted. During the evaluation, the Current/Past Experience Graph identified 27 episodes of disease reports that exceeded historical limits. Of these, 14 (52 percent) represented outbreaks. None were detectable by analysis of aggregate national surveillance data. Five outbreaks known to state health department officials were not identified by the method because of increased disease activity during the baseline period or lack of timely provisional reporting of outbreak-related cases.

What are the mechanics of operation?

For any analytic method to be useful, it must be easily implemented in the routine work of the practicing epidemiologist. During this evaluation, an epidemiologist routinely evaluated each aberration produced by the Current/Past Experience Graph, analyzing state distributions and communicating with each Centers for Disease Control and Prevention program area responsible for the control of the condition. Additional information was provided by epidemiologists in state health departments. Investigation was based on this evidence in addition to that provided by other methods of analysis. Development is ongoing to provide software to produce the Current/Past Experience Graph for use by state health departments.

We have developed a method for analyzing public health surveillance data that demonstrates high sensitivity and positive predictive value for some nationally notifiable diseases. Statistical evaluation shows that the method performs well under a wide variety of conditions. Epidemiologic assessment shows high sensitivity and positive predictive value.

Since this study was conducted, the method has remained useful in national surveillance. Recent increases beyond historical limits in reporting of aseptic meningitis reflected increased disease activity primarily in the northeastern United States (22). Increases in animal rabies reports were due to increased reporting of raccoon rabies in mid-Atlantic and northeastern States (unpublished data). Although not applicable to all types of surveillance questions, the method is useful for identifying conditions that require further investigation and for providing sensible solutions from imperfect data to facilitate public health action.

REFERENCES

1. Lilienfeld AM, Lilienfeld DE. Foundations of epidemiology. 2nd ed. New York, NY: Oxford University Press, 1980.
2. Thacker SB, Berkelman RL, Stroup DF. The science of public health surveillance. J Public Health Policy 1989;10:187–203.
3. Serfling RE. Methods for current statistical analysis of excess pneumonia-influenza deaths. Public Health Rep 1963;78:494–506.
4. Thacker SB, Choi K, Brachman PS. The surveillance of infectious diseases. JAMA 1983;249: 1181–5.
5. Centers for Disease Control. Summary of notifiable diseases, United States, 1990. MMWR Morb Mortal Wkly Rep 1990;39:1–61.
6. Stroup DF, Williamson GD, Herndon JL. Detection of aberrations in the occurrence of notifiable diseases surveillance data. Stat Med 1989;8: 323–32.
7. Centers for Disease Control. Proposed changes in format for presentation of notifiable disease report data. MMWR Morb Mortal Wkly Rep 1989;38: 805–9.
8. Stroup DF, Thacker SB, Herndon JL. Application of multiple time series analysis to the estimation of pneumonia and influenza mortality by age, 1962–1983. Stat Med 1988;7:1045–59.
9. Morgenstern H, Greenland S. Graphing ratio measures of effect. J Clin Epidemiol 1990;43:539–42.
10. Kafadar K, Stroup DF. Analysis of aberrations in public health surveillance data: estimating variances on correlated samples. Stat Med 1992;11: 1551–68.
11. Efron B, Tibshirani R. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Stat Sci 1986;1: 54–77.
12. Bose A. Edgeworth correction by bootstrap in autoregressions. Ann Stat 1988;16:1709–22.
13. Liu RY. Bootstrap under some non-i.i.d. models. Ann Stat 1988;16:1696–709.
14. Mosteller F, Tukey JW. Data analysis and regression: a second course in statistics. Reading, MA: Addison-Wesley, 1977.
15. Künsch HR. The jackknife and the bootstrap for general stationary observations. Ann Stat 1989:17: 1217–41.
16. Wharton ME, Price W, Hoesly F, et al. Evaluation of a method for detecting outbreaks of diseases in six states. Am J Prev Med 1993;9:45–9.
17. Centers for Disease Control. Increase in rubella and congenital rubella syndrome—United States, 1988–1990. MMWR Morb Mortal Wkly Rep 1991; 40:93–9.
18. Sacks JJ. Utilization of case definitions and laboratory reporting in the surveillance of notifiable communicable diseases in the United Sates. Am J Public Health 1985;75:1420–2.
19. Thacker SB, Berkelman RL. Public health surveillance in the United States. Epidemiol Rev 1988; 10:164–90.
20. Cleveland WS, Devlin SJ. Calendar effects in monthly time series: modeling and adjustment. J Am Stat Assoc 1982;77:520–8.
21. Wharton M, Price W, Hoesly F, et al. Evaluation of a method for detecting outbreaks of diseases in six states. Am J Prev Med 1993;9:45–9.
22. Centers for Disease Control. Aseptic meningitis—New York State and United States, weeks 1–36, 1991. MMWR Morb Mortal Wkly Rep 1991;40: 773–5.