

## A MONITORING SYSTEM FOR DETECTING ABERRATIONS IN PUBLIC HEALTH SURVEILLANCE REPORTS<sup>†</sup>

G. DAVID WILLIAMSON<sup>1\*</sup> AND GINNER WEATHERBY HUDSON<sup>2</sup>

<sup>1</sup>*Epidemiology Program Office, MS K73, Centers for Disease Control and Prevention, 4770 Buford Highway, Atlanta, GA 30341, U.S.A.*

<sup>2</sup>*Drexel University, College of Business and Administration, Matheson Hall, 32nd and Market Streets, Philadelphia, PA 19104, U.S.A.*

### SUMMARY

Routine analysis of public health surveillance data to detect departures from historical patterns of disease frequency is required to enable timely public health responses to decrease unnecessary morbidity and mortality. We describe a monitoring system incorporating statistical ‘flags’ identifying unusually large increases (or decreases) in disease reports compared to the number of cases expected. The two-stage monitoring system consists of univariate Box–Jenkins models and subsequent tracking signals from several statistical process control charts. The analyses are illustrated on 1980–1995 national notifiable disease data reported weekly to the Centers for Disease Control and Prevention (CDC) by state health departments and published in CDC’s *Morbidity and Mortality Weekly Report*. Published in 1999 by John Wiley & Sons, Ltd. This article is a U.S. Government work and is in the public domain in the United States.

### INTRODUCTION

Public health surveillance provides the foundation for much of effective epidemiologic and public health practice. It is the ongoing collection, analysis, interpretation and dissemination of outcome-specific information for use in planning, implementing and evaluating public health practice.<sup>1,2</sup> Data collected from public health surveillance systems can provide important clues to the aetiology of a disease and assist in identification of important risk factors, onset of epidemics and detection of unusual observations in reports of infectious diseases and other conditions, thus facilitating early public health response to minimize undue morbidity and mortality.<sup>3–5</sup>

There is a long distinguished history in modelling public health data to provide greater insight into aetiology, spread, prediction and control of diseases. One such early effort was by William Farr when, in 1840, he fit a normal curve to deaths from smallpox in hopes of discovering why epidemics appear and disappear.<sup>6</sup> Other epidemiologic and statistical analyses have focused on modelling disease incidence and prevalence, geographical distribution and spread of disease, and on forecasting disease counts and associated health care needs.<sup>7–17</sup> Models have been developed to detect time and/or spatial clusters of disease and, recently, to smooth rates in small area estimation problems (that is, to overcome statistical issues occurring when the unit of analysis is associated with a geographic area which is small relative to the area spanned by a set of

\* Correspondence to: G. David Williamson, Epidemiology Program Office, MS K73, Centers for Disease Control and Prevention, 4770 Buford Highway, Atlanta, GA 30341, U.S.A. E-mail: dxw2@cdc.gov

<sup>†</sup>This article is a U.S. Government work and is in the public domain in the United States.

contiguous geographic areas analysed).<sup>18–23</sup> Because some epidemiologic and, in particular, public health surveillance data are collected in time sequence at regular intervals in an ongoing manner, these data often exhibit correlation, non-stationarity (in the mean and/or variance) and seasonality, characteristics which statistical time series modelling is especially suited to accommodate.<sup>24</sup>

Of critical importance to public health practitioners is an ability to detect quickly substantial changes in disease and other consequential epidemiologic data series, thus facilitating timely public health response to mitigate morbidity and mortality.<sup>25–27</sup> Because of the special features of public health surveillance data as listed above, and including that the data are not usually generated from a random sample, detection of changes in public health data presents an analytic challenge. However, particularly in the last 15 years, several research efforts have described statistical and epidemiologic methods to detect substantial changes in public health data series, including timely signalling of the onset of epidemics or identification of aberrations (that is, statistically significant departures in the occurrence of a health event from what is expected based on the historical incidence of the event) in the data.<sup>27–36</sup> These methods include a bar graph based on the ratio of current to historical data, extensions to the linear dynamic model and applications of the Kalman filter and probability index function. Additionally, Box–Jenkins time series models have been applied to public health forecasting problems in the past<sup>13,37–41</sup> and statistical process control (SPC) methods have been discussed as evaluation tools for public health surveillance,<sup>42</sup> although we are unaware of any previous literature describing a combination of the two methods for a public health monitoring system.

The objective of this research was to develop a monitoring system to (i) detect aberrations in reported data on disease incidence, and (ii) provide a signal to alert public health practitioners to undertake timely public health action. Here we develop and introduce a two-stage monitoring system comprised of two traditional time series techniques, Box–Jenkins autoregressive, integrated, moving average (ARIMA) models and SPC charts, for the detection and signalling of aberrations in public health data. A substantial difference between other techniques and our monitoring system is that we combine the strengths of Box–Jenkins and SPC methods, taking into account the order of the data in time as well as explicitly incorporating methods which recognize and account for the correlation among observations which occur due to the influence of the same (epidemiologic) factors at adjacent or nearby time periods (referred to as autocorrelation in the time series literature).

In this paper, we describe our innovative modelling approach to detecting aberrations and apply the methods to United States disease report series. We consider issues which arise when employing the monitoring system and discuss further directions to pursue for enhancing our ability to detect aberrations in public health surveillance data.

## METHODS

### Data

Analyses for this investigation were performed on data from the National Notifiable Diseases Surveillance System (NNDSS) of the Centers for Disease Control and Prevention (CDC). The NNDSS database consists of weekly reports of 52 diseases (as of 1 January 1996) designated by the Council of State and Territorial Epidemiologists as nationally notifiable and approved by the state and territorial health departments for reporting to CDC.<sup>43</sup> A notable strength of the

NNDS is its timeliness. Reports of cases of diseases are aggregated weekly and sent by each state health department to CDC for dissemination in CDC's *Morbidity and Mortality Weekly Report*. Data analysed here are provisional (that is, do not include updates from states as additional information becomes available) to allow for the most timely monitoring of disease series and any necessary subsequent public health action.

Data analyses were performed in two phases, the first beginning in 1990 and the second in 1995. For the original analysis, 17 diseases were chosen to represent a diversity of disease and demographic characteristics, including disease aetiology, seasonality and incidence, population distribution and geography. These diseases were acquired immunodeficiency syndrome (AIDS), aseptic meningitis, encephalitis, gonorrhoea, hepatitis type A, hepatitis type B, hepatitis non-A and non-B, legionellosis, malaria, measles, meningococcal infections, mumps, pertussis, rubella, syphilis, tuberculosis and typhus fever (tickborne). In addition to investigating application of our methods on the national data for these 17 diseases, we applied our modelling strategy and control charts to state data for gonorrhoea, hepatitis type A, measles, meningococcal infections and typhus fever (tickborne). We applied the methods to data from each of three states for hepatitis type A, meningococcal infections and typhus fever (tickborne), and to data from each of four states for the other two diseases. These diseases and states were chosen to represent a wide array of disease and demographic characteristics. This first phase included data from week one 1980 to week 52 1989. The second analysis phase augmented the original data with disease reports from week one 1990 to week 19 1995 and has focused to date only on the national disease series for hepatitis type A. Approximately every six years there is a 53rd week of data included in the disease series. For these years, we average the number of reported cases for weeks 52 and 53, creating a new value for week 52 and ensuring consistency in the number of weeks each year for analysis.

### Analysis Strategy

Each of our disease report series is a time-dependent phenomenon which is affected by various factors, some known and some unknown. Because of the complexity and unknown nature of some of the factors affecting disease reports, stochastic rather than deterministic models are used to better represent and forecast numbers of reported cases of disease. Many disease series contain trend, seasonal and cyclical components which result in substantial autocorrelation. We employ the Box-Jenkins modelling strategy on the surveillance data because these techniques are designed to properly handle autocorrelation issues.<sup>24</sup> If the ARIMA Box-Jenkins time series models provide an adequate fit to the data, they will produce one-step-ahead minimum variance forecasts of the process. These forecasts are the expected number of reported cases of disease and are based on the historical provisional data. The associated forecast errors (that is, the differences between the model forecasts and the observations) should then be approximately independently and identically distributed (IID).

Statistical process control charts are among the most prevalent and valid methods for monitoring time series data, but their use usually requires observations to be IID random variables when the process is in statistical control.<sup>44</sup> If assumptions for applying Box-Jenkins methods are met and adequate ARIMA models developed, then the subsequent forecast errors from the models can be monitored by the SPC charts. Thus the objective of an appropriate monitoring system is not to track week-to-week changes in the number of reported cases, but rather to identify any substantial variation in the weekly expected number of reported cases as

determined from historical patterns. The control charts employ statistical limits to generate 'flags' identifying deviations from historical data patterns and from the underlying stochastic process generating the observations (that is, the charts detect aberrations). These aberrations signal a potential change in the pattern of incidence of that health event or in the associated reporting procedures, thus facilitating a timely public health response.

Thus we have developed and present here a two-stage monitoring system for public health surveillance data based on the successful integration of (i) ARIMA time series models providing dynamic forecasting of future expected disease reports and (ii) SPC methods for tracking the forecast errors from the ARIMA models. Similar approaches have been suggested for application in manufacturing and industrial settings.<sup>45,46</sup>

### *Preliminary analysis*

We produced several time plots for each of the national and state disease report series we analysed to better understand the historical data and guide the Box–Jenkins and SPC analyses. These graphics included (i) number of reported cases each week (for example, from week one 1980 to week 52 1989 for the first analysis phase) to visually portray the entire series of data; (ii) number of reported cases each week over the last two years of the period of analysis to more carefully depict recent patterns in historical data; (iii) number of reported cases each week plotted vertically by year (that is, each of the 52 weekly reports plotted in a vertical line above the 'year' label on the horizontal axis) and connected at the annual mean to investigate evidence of non-stationarity in the mean level and variability of weekly reports within each year of the series; and (iv) number of reported cases each week plotted vertically by week number, one to 52 (for example, for 1980–1989 analysis, each of the ten week-one reports, one for each year, are plotted in a vertical line above the 'week one' label on the horizontal axis, with data for the other 51 weeks plotted similarly) and connected at the weekly mean to show the range of values for the same week number across years, thus revealing any seasonality in the data.

### *Development of Box–Jenkins ARIMA time series models for forecasting*

In order to apply Box–Jenkins methods to time series data, the series must meet certain stationarity assumptions; the series must be stationary in both mean and variance across the modelling period. A series not stationary in the mean can be made so generally by calculating a differenced series (that is, by taking the differences in number of reported cases of disease between two specified reporting periods, generally successive ones). A series not stationary in the variance of its observations can many times be made so by transforming the data (for example, with a square root transformation).<sup>47,48</sup>

For those data series stationary in both the mean and variance, we applied Box–Jenkins techniques to develop forecasting models.<sup>24,47,48</sup> Often mathematical transformations and/or differencing was required to achieve stationarity. The Box–Jenkins approach to time series forecasting requires an adequate stochastic ARIMA model of the following form to describe the time series under study:

$$\Phi_p(B) \nabla^d z_t = \Theta_q(B) a_t$$

where  $\Phi_p(B) = (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)$  is an autoregressive polynomial of order  $p$ ,  $\nabla$  is the backward difference operator,  $d$  is the order of the first difference,  $z_t$  is the observation at time  $t$ ,  $\Theta_q(B) = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q)$  is a moving average polynomial of order  $q$ ,  $B$  is the

backshift operator, and  $a_t$  is a sequence of normally and independently distributed random 'shocks' with mean zero and constant variance.

To develop and select an appropriate ARIMA model for each series, extensive identification, estimation and diagnostic checking (that is, model evaluation) stages were completed. This work was accomplished using SAS/ETS<sup>R</sup> software.<sup>49</sup> The modelling procedure for each series began by examining the correlative structure of the data in the autocorrelation and partial autocorrelation functions to identify tentative models. Parameters were then estimated for each tentative model using disease report data for a selected period.

Model residuals, the difference in observed and model-estimated values for the historical period used in model estimation, and forecast errors were examined in the diagnostic checking stage, which was comprised of evaluating statistical fit and model forecasting capabilities. Forecast errors were determined by withholding a period of the most recent reports from the historical data used to estimate the ARIMA model parameters. A comparison of the model forecast with the actual reported cases that had been withheld could then be made and the corresponding forecast errors (actual minus forecast values) calculated. The adequacy of the statistical fit was determined by the statistical significance of any autoregressive (AR) or moving average (MA) terms in the model and by analysis of the model residuals, which should be basically free of autocorrelation and patterns. Model performance for the estimation period and the forecast period was quantified on the residuals and forecast errors, respectively, by calculating various statistics, including mean absolute deviation (MAD) and root mean squared error (RMSE).

In both analysis phases of this work, beginning in 1990 and again in 1995, the historical data were divided into two periods, one for model estimation and one for forecasting, model evaluation and testing. For the 1990 analysis phase, we developed a model for the 1980–1987 historical data and forecasted weekly reported cases for all 52 weeks in 1988, the forecast for the 52nd week of 1988 being termed a 52-week-ahead forecast. Then, to check the robustness of the form of the ARIMA model to changes in the data (that is, to inclusion of additional observations), using the same model form as for the 1988 forecasts, we re-estimated the model coefficients with 1980–1988 historical data and forecasted all 52 weekly values for 1989, ultimately producing a 52-week-ahead forecast for the last week of 1989. Within production of the 1989 forecasts, we did not incorporate the more up-to-date 1989 disease reports and there was no redevelopment of model form or re-estimation of model coefficients after each weekly forecast was calculated.

For the 1995 analysis phase, we developed an ARIMA model for 1980–1993 data, then forecasted weekly values for 1994 and the first 19 weeks of 1995 in two ways. First we developed a model for the 1980–1993 data and forecasted weekly values for all 52 weeks of 1994 and the first 19 weeks of 1995, ultimately producing a 71-week-ahead forecast. In this first approach, we did not incorporate the more up-to-date 1994–1995 disease reports nor was there redevelopment of model form or re-estimation of model coefficients after each weekly forecast was calculated. Secondly, we developed one-week-ahead, rolling forecasts for each of the 71 weeks of 1994 and 1995 by retaining the same model form at each forecasting step, but re-estimating the model coefficients after the addition of each week's newly reported cases from 1994 and first 19 weeks of 1995 to incorporate as much data as were available in the estimation step, and thus produce a more valid forecast. This second approach was consistent with recommendations that Box–Jenkins models produce the most valid results when forecasting one-step-ahead values.<sup>24</sup> This latter strategy was also implemented to simulate how the monitoring system might be

utilized by CDC and state health departments and to test the robustness of the model coefficients to the inclusion of additional data each week by allowing comparison of results from the two manners in which 1994–1995 data were forecasted.

#### *Use of SPC methods and control charts for tracking signals*

Statistical process control charts are a major category of methods for creating tracking signals. These methods, dating back to Walter Shewhart's work, employ collecting data sequentially in time and plotting those observations or functions of the observations on control charts.<sup>50</sup> In the basic Shewhart-type chart, a plotted point is compared to predetermined control limits, and if the point falls beyond these critical boundaries, it is a signal that the process is statistically out of control, or that a statistical aberration has been identified. Although calculation of control chart statistics is relatively easy, it is sometimes difficult to determine the most effective control charts and appropriate control limits for the specific monitoring problem.<sup>44</sup>

In the proposed system, we considered the following three types of control charts: (i) Shewhart; (ii) moving average; and (iii) exponentially weighted moving average (EWMA). These charts were implemented using SAS/QC<sup>R</sup> software.<sup>51</sup>

Upper and lower control limits in the Shewhart control chart are typically set at  $\pm 3$  standard deviations from the overall average level. The Shewhart chart detects large deviations (1.5 to 2.0 standard deviations or greater) from the previous stable pattern very quickly, but is not as effective in detecting smaller shifts. This chart considers only the last plotted individual point, thus ignoring information about the process in previous observations.

The moving average control chart is similar in application to the Shewhart chart, but is more effective in detecting small process shifts. For this control chart the moving average of span  $w$ , the average of the last  $w$  points, is plotted rather than the individual point.

The EWMA control chart, first presented as the 'geometric moving average chart', combines historical data to give less weight to data as they get older:

$$\hat{y}_t = \lambda y_t + (1 - \lambda) \hat{y}_{t-1}$$

where  $\hat{y}_t$  is the statistic at time  $t$  (the new EWMA),  $\hat{y}_{t-1}$  is the statistic at time  $t - 1$  (the old EWMA),  $\lambda$  is the EWMA weighting parameter and  $y_t$  is the observed value at time  $t$  (the new observation).<sup>52</sup> The EWMA control chart can be designed to resemble the performance of the Shewhart control chart by the selection of  $\lambda$  in the interval (0, 1). The greater the value of  $\lambda$ , the smaller is the influence of the data in the more distant past. The value of  $\lambda$  can be determined subjectively,<sup>53</sup> but is commonly set at approximately 0.2 in EWMA control chart applications. Since the EWMA is a weighted average of observations, it is less sensitive to the normality assumption and, therefore, provides more flexibility in its application to monitoring problems.

For the 1990 analysis phase, we implemented the SPC charts to monitor and identify aberrations in the forecast errors for 1989 data. For the phase two analysis, we implemented the monitoring system on the 71 weeks of forecast errors available for 1994–1995.

## RESULTS

### **Box–Jenkins Time Series Models**

Most of the national and state disease series of reported cases were non-stationary in the mean and/or variance, so required data transformations and/or differencing to attempt to achieve those

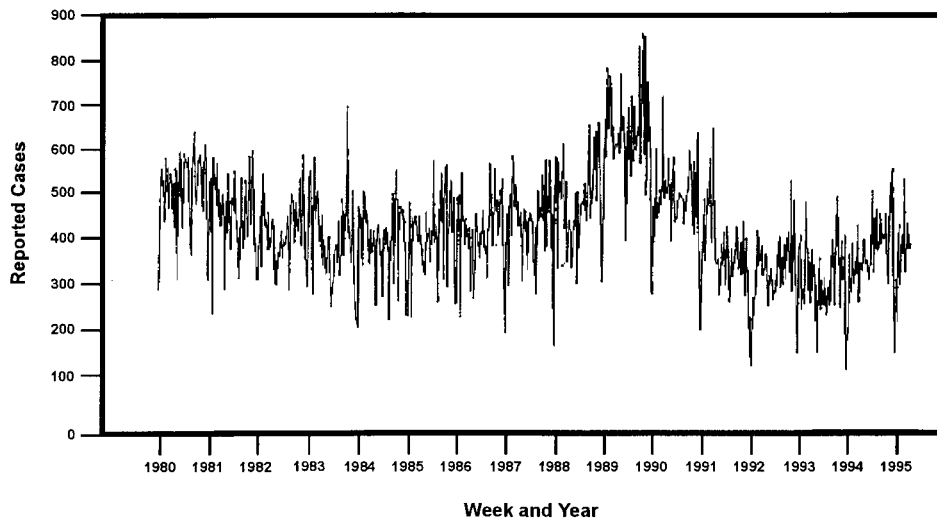


Figure 1. Number of reported cases of hepatitis A, by week, United States, 1980–1995 (provisional data 5 January 1980–13 May 1995)

stationarity requirements for applying Box–Jenkins time series methods. When series with non-stationary means were encountered, we employed differencing and, when non-stationary variances were observed, we used logarithm and square root transformations to increase stability and, hopefully, meet stationarity assumptions. Of the 17 national disease report series, we were able to meet stationarity assumptions and develop adequate Box–Jenkins ARIMA models for hepatitis type A, hepatitis type B, hepatitis non-A and non-B, legionellosis, malaria, meningococcal infections and tuberculosis. Of the state series, we successfully achieved stationarity and developed ARIMA models for all three state series for hepatitis type A and for one state for typhus fever (tickborne). Because of the modelling success at the national and state level for hepatitis type A, we focused analyses on those series and, subsequently, focus reporting here on hepatitis A, although, when meaningful, we make reference to results for other series.

For the hepatitis type A national disease report series (Figure 1), we performed first differencing of order one and 52 to achieve stationarity in the mean for the 1980–1988 data (Figure 2 depicts the first differences of order one) to appropriately apply Box–Jenkins methods for forecasting disease reports for 1989. The final model selected for this series, based on evaluation of statistical fit and model forecasting capabilities, was

$$(1 - 0.22B^{52})(1 - B)(1 - B^{52})z_t = (1 - 0.90B)(1 - 0.82B^{52})a_t.$$

The plot of observed versus model values (Figure 3) indicates the model provides an excellent fit to the data. This model form was the same as and the model coefficients were extremely similar to those of the model developed for forecasting 1988 disease report data based on the 1980–1987 data:

$$(1 - 0.21B^{52})(1 - B)(1 - B^{52})z_t = (1 - 0.92B)(1 - 0.78B^{52})a_t.$$

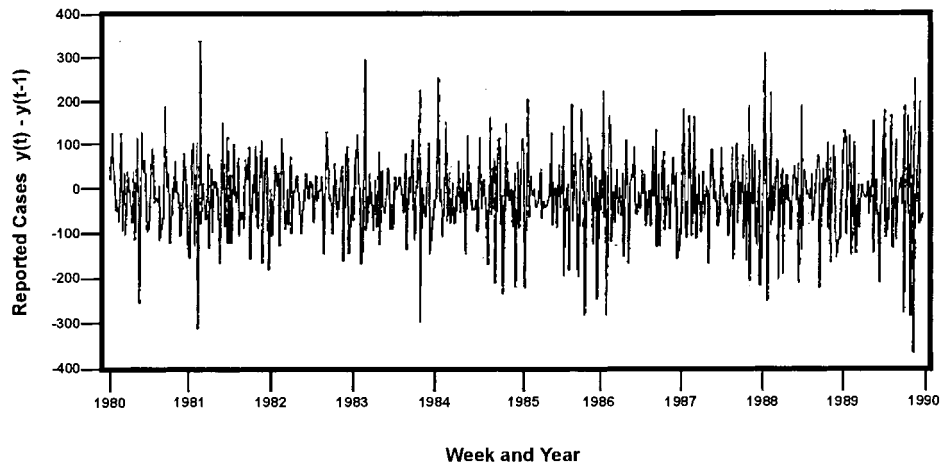


Figure 2. First differences in the number of reported cases of hepatitis A, United States, 1980–1989 (the first difference represents the week-to-week changes in the reported number of cases,  $y(t)$ )

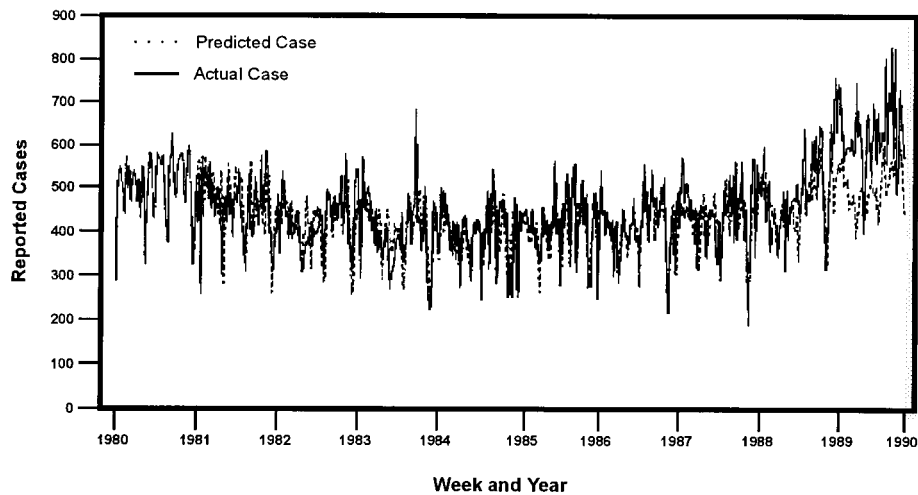


Figure 3. Predicted and actual number of reported cases of hepatitis A, by week, United States, 1980–1989 (the 1981–1988 predicted numbers are model estimates, the 1989 predicted numbers are forecasts and the actual number of cases are provisional data from 5 January 1980 to 30 December 1989)

The final model for hepatitis A for 1980–1993 national data, used to develop the monitoring system for 1994–1995 disease report data, was estimated using the square root of the reported cases data:

$$(1 - B)(1 - B^{52})z_t = (1 - 0.85B)(1 - 0.79B^{52})a_t.$$

This model was used to forecast weekly reports for 1994 and the first 19 weeks of 1995 in the two ways given above: (i) forecasted weekly values for all 71 weeks of 1994–1995 in the analysis with

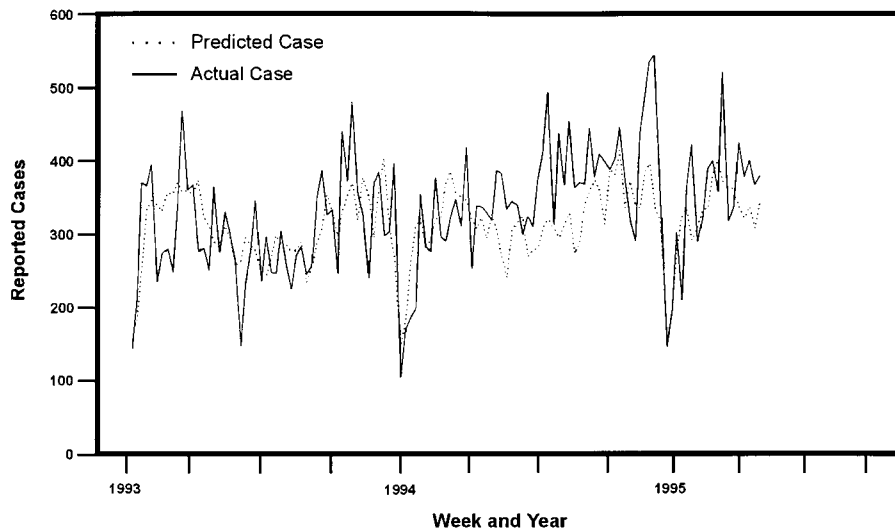


Figure 4. Predicted and actual number of reported cases of hepatitis A, by week, United States, 1993–1995 (the 1993 predicted numbers are model estimates, the 1994–1995 predicted numbers are 71-week-ahead forecasts and the actual number of cases are provisional data from 9 January 1993 to 13 May 1995)

no incorporation of 1994–1995 disease reports or redevelopment of the model form or re-estimation of model coefficients after each weekly forecast was calculated (Figure 4) and (ii) one-week-ahead, rolling forecasts for each of the 71 weeks using the same model form but re-estimating the model coefficients after the addition of each week's newly reported cases from 1994–1995 (Figure 5).

Although we modelled the hepatitis A disease report series for all three states on 1980–1988 data, the forms of the models were different from each other and from that for the national series discussed above.

### Statistical Process Control Methods

Shewhart and moving average (with span of size two) control charts developed on national hepatitis type A values forecasted for 1989 (phase one analysis) identified several statistically high values in the data (Figures 6 and 7). The EWMA chart detected a statistically significant deviation from the forecasted or expected values, thus indicating the likelihood of an increasing trend throughout the series (Figure 8).

Of the other six national disease report series for which we developed models and applied the SPC charts in the phase one analysis (through 1989), only for hepatitis B did more than one chart identify any significantly high aberrations (Shewhart and moving average charts identified a few high values close in time and the EWMA chart identified statistical aberrations indicating the likelihood of an increasing trend). For malaria the moving average chart identified a single statistically high aberration. The monitoring system detected no significantly high values for any of the hepatitis non-A and non-B, legionellosis, meningococcal infections and tuberculosis forecasted data. In phase two of the analysis (to week 19 of 1995), none of the three SPC charts

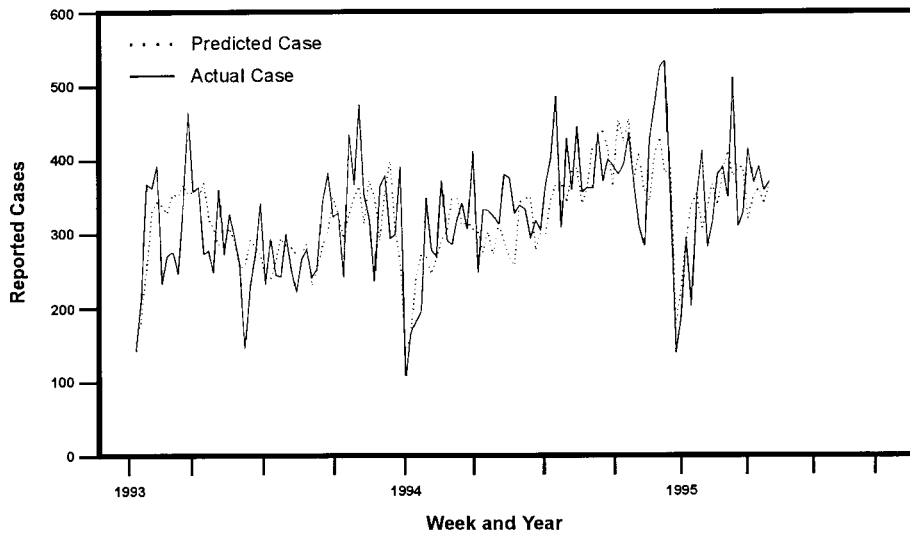


Figure 5. Predicted and actual number of reported cases of hepatitis A, by week, United States, 1993–1995 (the 1993 predicted numbers are model estimates, the 1994–1995 predicted numbers are rolling forecasts and the actual number of cases are provisional data from 9 January 1993 to 13 May 1995)

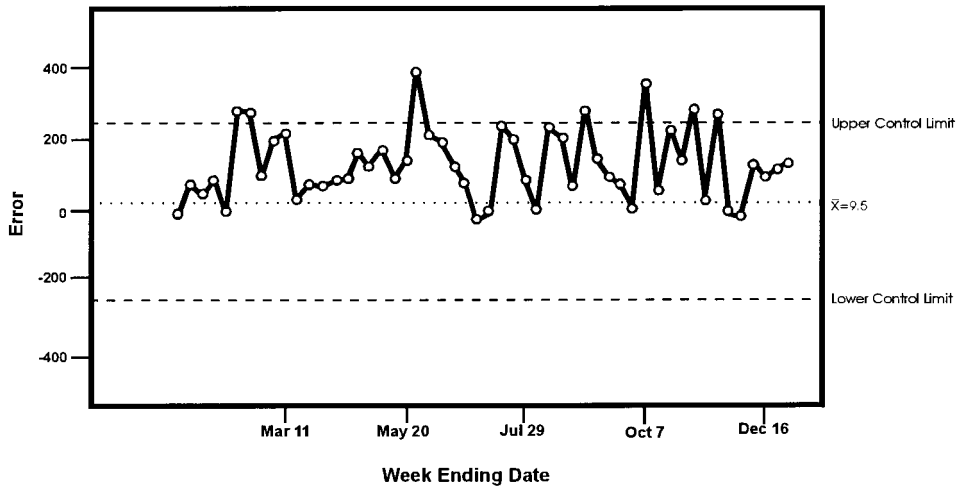


Figure 6. Shewhart control chart for the number of reported cases of hepatitis A, by week, United States, 1989

identified any significantly high values for the 1994–1995 forecasted national hepatitis A disease report series.

For 1989 hepatitis A forecasted data, monitoring results varied by state with no evident relationships for statistically high values among the three states or with the national disease monitoring results. High aberrations were detected by at least one control chart for each of the

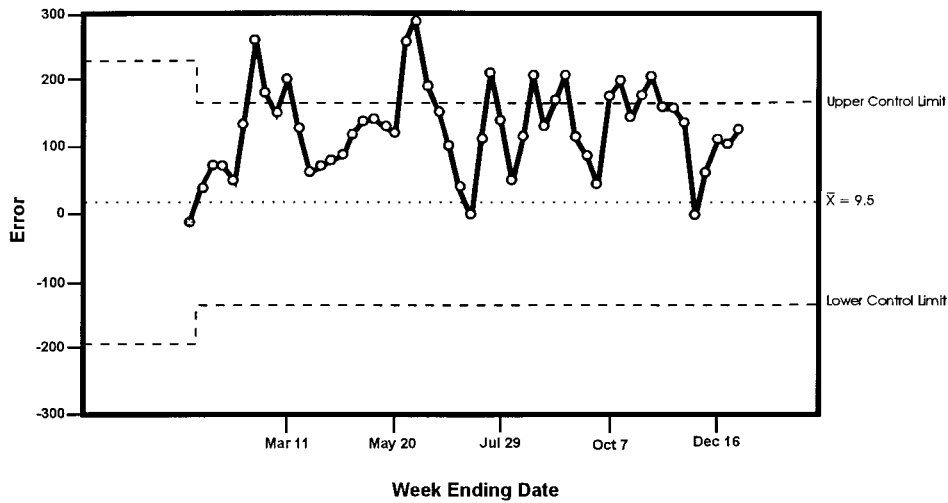


Figure 7. Moving average control chart for the number of reported cases of hepatitis A, by week, United States, 1989

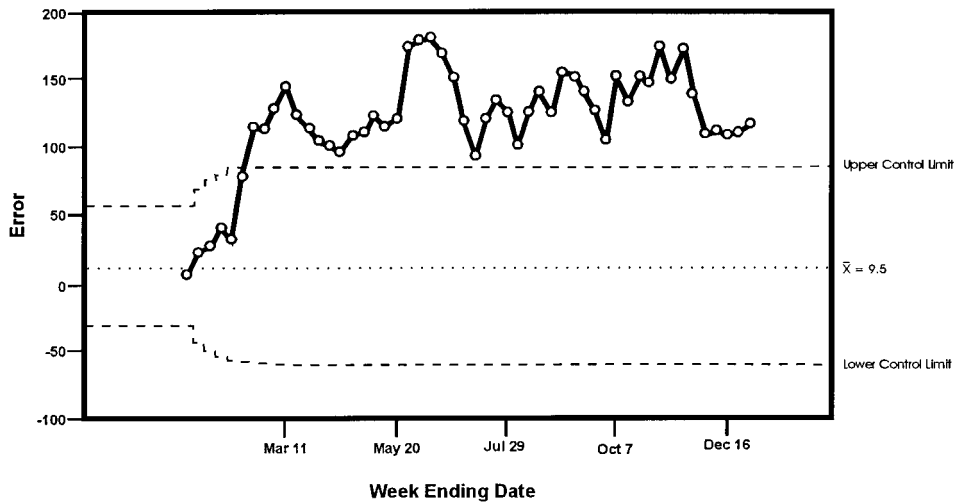


Figure 8. Exponentially weighted moving average control chart for the number of reported cases of hepatitis A, by week, United States, 1989

three states, with all three charts (Shewhart, moving average and EWMA) signalling high values for one state.

### DISCUSSION

A critical aspect of public health is to monitor incidence, prevalence and patterns of occurrence and spread of diseases and other health events, and to report those to the public and the public

health community so that appropriate actions are taken. In part because communication technology continues to improve and funding sources require increased justification for scientific research, it is incumbent for public health practitioners to use all available data and analysis tools to make timely, appropriate decisions. We introduce here a monitoring system which holds promise for aiding in identification of unusual observations in public health surveillance data by providing early warning signals.

Because our research focus was on viability of quantitative tools rather than on providing in-depth insight into aetiology or risk factors of disease, we limit the discussion here to use and understanding of the methods, and provide no programmatic guidance regarding disease occurrence, prevention and control. However, the scope and diversity of diseases within the NNDSS provides a substantive arena for evaluation of the monitoring system and has guided us in establishing an ambitious research plan for the future.

We applied our analyses to reported number of cases of disease rather than to reported rates of disease in part because the NNDSS does not include population data, a characteristic shared by most public health surveillance systems. Additionally, because we were investigating potential changes in patterns of disease reports at different times for the same reporting areas, we would not expect dramatic changes in population size for the areas and, thus, analyses of rates would not provide substantial differences in results. Should the focus of analyses shift to comparison among reporting areas with different population sizes, then analyses of rates would be more appropriate. Our methods, however, could readily be applied to rates.

We successfully developed Box–Jenkins ARIMA models for seven of the 17 national disease report series and four of the 17 state disease series we investigated. For those series we were unable to model, the deviations were too great to achieve stationarity and, in some cases, data were too sparse for much of the reporting period to apply the Box–Jenkins methods. This was especially true for the state series when there was a substantial increase in amounts of missing data and low or no reported cases of disease. Another hindrance to the modelling process was the lack of pattern in the disease report series, such as that associated with clustering of reports of disease occurrence at random, non-seasonal times.

A strength of these modelling techniques was the stability demonstrated as additional data became available over two years; recall the form and coefficients of the model were very similar for the 1988 and 1989 forecasting models for hepatitis A. Models developed for the other six national disease series demonstrated this same similarity and stability. With addition of 1990–1993 national hepatitis A data in the phase two analysis, we found the original model form (based on 1980–1989 data) still produced good forecasts, but that a model of substantially different form provided better performance, thus indicating the need to periodically review and change the model form to maintain the best (or perhaps valid) forecasting model and maintain a reliable monitoring system for detecting aberrations.

The greatest utility for Box–Jenkins models comes with one-step-ahead forecasts because forecast errors for longer lead times will in general be correlated. A major advancement in our system occurred when we implemented this one-step-ahead, rolling forecast method in the phase two analysis. Figures 4 and 5 allow comparison of the two forecast methods and show how, when one-week-ahead rolling forecasts are utilized (Figure 5), the forecasts for 1994–1995 more closely mirror the observed values. However, because the model is readjusted with the addition of each new observation when using rolling forecasts, the monitoring system (which would then be implemented on forecasted errors from a continually readjusting model) may have a tendency to miss picking up trends in the disease series, an issue which needs to be evaluated in future work.

A combination or blend of forecasting strategies with specific decision rules may provide the most effective monitoring system in practice.

Because the national disease series is comprised of reports from 50 states, it was not surprising that the forms of the models developed for hepatitis A for the three states were different from the form for the national model. Also, because the three states used in the hepatitis A analysis were geographically dispersed, it was not unexpected that model forms for the states also differed from each other. Future research should include analysis on data from contiguous states to determine whether similar models are developed for those states with similar disease report patterns.

The Shewhart and moving average control charts, both effective at identifying relatively large shifts from the overall average level, signalled high values at similar times for the 1989 hepatitis A forecasted data (Figures 6 and 7). One of the strengths of our monitoring system is employing several SPC charts to detect different types and sizes of aberrations (for example, the EWMA chart can detect small shifts and gradual trends in the data and, thus, complements the capabilities of Shewhart and moving average charts to detect larger, acute changes in the mean of the series). The cumulative sum, or CUSUM, SPC chart is designed to identify smaller sustained shifts in a process and, thus, should yield analysis results similar to those when employing the EWMA chart.<sup>54</sup> We have performed initial work incorporating the CUSUM chart in the monitoring system but, because the research is preliminary, have excluded those results here. VanBrackle and Williamson, in other work presented in this issue, demonstrate how the monitoring system with several control charts detects spike, step and trend changes of varying sizes in the disease series.<sup>36</sup> Their research also includes how correlated data affect capabilities of the monitoring system to detect aberrations. Some research has suggested the use of a combined Shewhart–EWMA control procedure, which should be effective against both large and small shifts.<sup>44</sup> In this combined approach, both the EWMA statistic and the individual forecast error could be plotted on the same chart with the two sets of control limits. However, there is still much work to be done in application of SPC charts within our monitoring system and evaluation of subsequent results, including further application of the CUSUM chart.

There are other important directions related to the problem of detecting aberrations which need to be investigated. Additional methods, such as neural network techniques and fractional differencing, smoothness priors and time-space time series methods, should be applied to the NNDSS data, evaluated, and compared with results from our monitoring system work. Monteiro *et al.* present a comparison of dynamic linear models and ARIMA models on NNDSS data.<sup>41</sup> The effects of multiple testing should be evaluated and perhaps accounted for in the monitoring system. Further work should be developed to examine earliest detection of changes in patterns of disease report data (to identify beginning stages of an epidemic), such as with change point models.<sup>41</sup>

A critical aspect to evaluation of any method is to validate its usefulness in the intended setting. Although we had periodic discussions with disease experts during this research, we need to perform an epidemiologic validation to determine sensitivity and specificity of our monitoring system, realizing that no monitoring system will be able to discern between a disease-related increase and one caused by clerical error, batch reporting or statistical anomaly. This evaluation should be done on the local or state level, from where the data originate and public health responses to substantial increases in reported cases of disease are likely to come.

To affect utilization of analytic methods in rapid response situations, as with potential epidemics, it is crucial that public health practitioners know and understand those methods and have an easy, timely, inexpensive way to implement them. CDC's Statistical Software for Public

Health Surveillance, which provides several methods for analysing surveillance data, including the Box–Jenkins time series techniques, is available through the Internet, but at this time lacks the SPC charts needed to develop the monitoring system described here.<sup>55</sup>

Although this research has focused on our monitoring system for detecting aberrations, it has reinforced thoughts that, because of the breadth and diversity of diseases and the factors affecting them, there is no single method which can be universally applied to public health surveillance data to identify an unusually high number of cases of disease or health event. The monitoring system is a new approach which maximizes strength from two time series methods, and it is an automated, flexible one which shows promise of assisting the public health community in these efforts. However, there remain numerous opportunities for development, application and evaluation of quantitative methods to aid in identifying outbreaks, sentinel public health events and aberrations in disease data, and, thus, facilitate timely actions to decrease unnecessary morbidity and mortality.

#### REFERENCES

1. Thacker, S. B. and Berkelman, R. L. 'Public health surveillance in the United States', *Epidemiologic Review*, **10**, 164–190 (1988).
2. Teutsch, S. M. and Churchill, R. E. (eds). *Principles and Practice of Public Health Surveillance*, Oxford University Press, New York, 1994.
3. Thacker, S. B., Berkelman, R. L. and Stroup, D. F. 'The science of public health surveillance', *Journal of Public Health Policy*, **10**, 187–203 (1989).
4. Noah, N. D. 'Cyclical patterns and predictability in infection', *Epidemiologic Infections*, **102**, 175–190 (1989).
5. 'National conference on clustering of health events', *American Journal of Epidemiology (Supplement)*, **132**, S1–S200 (1990).
6. Farr, W. *Progress of epidemics. Second Report of the Registrar General of England and Wales*, His Majesty's Stationery Office, London, 1840, pp. 91–98.
7. Greenwood, M. 'On the statistical measure of infectiveness', *Journal of Hygiene*, **31**, 336–351 (1931).
8. Abbey, J. 'An examination of the Reed-Frost theory of epidemics', *Human Biology*, **24**, 201–203 (1952).
9. Serfling, R. E. 'Methods for current statistical analysis of excess pneumonia-influenza deaths', *Public Health Reports*, **78**, 494–506 (1963).
10. Lui, K.-J. and Kendal, A. P. 'Impact of influenza epidemics on mortality in the United States from October 1972 to May 1985', *American Journal of Public Health*, **77**, 712–716 (1987).
11. Brookmeyer, R. and Liao, J. 'Statistical modelling of the AIDS epidemic for forecasting health care needs', *Biometrics*, **46**, 1151–1163 (1990).
12. Cliff, A. D., Haggett, P. and Stroup, D. F. 'The geographic structure of measles epidemics in the northeastern United States', *American Journal of Epidemiology*, **136**, 592–602 (1992).
13. Cliff, A. D. and Haggett, P. 'Statistical modelling of measles and influenza outbreaks', *Statistical Methods in Medical Research*, **2**, 43–73 (1993).
14. Sprenger, M. J. W., Mulder, P. G. H., Beyer, W. E. P., Strik, R. V. and Masurel, N. 'Impact of influenza on mortality in relation to age and underlying disease, 1967–1989', *International Journal of Epidemiology*, **22**, 334–340 (1993).
15. Cliff, A. D., Haggett, P., Smallman-Raynor, M. R., Stroup, D. F. and Williamson, G. D. 'The application of multidimensional scaling to epidemiologic data', *Statistical Methods in Medical Research*, **4**, 102–123 (1995).
16. Simonsen, L., Clarke, M. J., Stroup, D. F., Williamson, G. D., Arden, N. H. and Cox, N. J. 'A method for timely assessment of influenza-associated mortality in the United States', *Epidemiology*, **8**, 390–395 (1997).
17. Simonsen, L., Clarke, M. J., Williamson, G. D., Stroup, D. F., Arden, N. H. and Schonberger, L. B. 'The impact of influenza epidemics on mortality: introducing a severity index', *American Journal of Public Health*, **87**, 1944–1950 (1997).

18. Ingram, D. D. and Williamson, G. D. 'Statistical and epidemiologic techniques. Summary of major tests available for statistical assessment of clustering, in Recommendations and Reports, July 27, 1990', *Morbidity and Mortality Weekly Report*, **39**, (RR-11), 17-23 (1990).
19. Besag, J. and Newell, J. 'The detection of clusters in rare diseases', *Journal of the Royal Statistical Society, Series A*, **154**, 143-155 (1991).
20. Marshall, R. J. 'Mapping disease and mortality rates using empirical Bayes estimators', *Applied Statistician*, **40**, 283-294 (1991).
21. Marshall, R. J. 'A review of methods for the statistical analysis of spatial patterns of disease', *Journal of the Royal Statistical Society, Series A*, **154**, 421-441 (1991).
22. Cressie, N. 'Smoothing regional maps using empirical Bayes predictions', *Geographical Analysis*, **24**, 75-95 (1992).
23. Elliott, P., Cuzick, J., English, D. and Stern, R. (eds). *Geographical and Environmental Epidemiology Methods for Small-Area Studies*, Oxford University Press, New York, 1992.
24. Box, G. E. P. and Jenkins, G. M. *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco, 1976.
25. Thacker, S. B. 'Historical development', in Teutsch, S. M. and Churchill, R. E. (eds), *Principles and Practice of Public Health Surveillance*, Oxford University Press, New York, 1994.
26. Cates, Jr., W. and Williamson, G. D. 'Descriptive epidemiology: Analyzing and interpreting surveillance data', in Teutsch, S. M. and Churchill, R. E. (eds), *Principles and Practice of Public Health Surveillance*, Oxford University Press, New York, 1994.
27. Stroup, D. F., Williamson, G. D., Herndon, J. L. and Karon, J. M. 'Detection of aberrations in the occurrence of notifiable diseases surveillance data', *Statistics in Medicine*, **8**, 323-329 (1989).
28. Smith, A. F. M. and West, M. 'Monitoring renal transplants: an application of the multiprocess Kalman filter', *Biometrics*, **39**, 867-878 (1983).
29. Centers for Disease Control and Prevention. 'Proposed changes in format for presentation of notifiable disease report data', *Morbidity and Mortality Weekly Report*, **38**, (47), 805-809 (1989).
30. Gordon, K. and Smith, A. F. M. 'Modeling and monitoring biomedical time series', *Journal of the American Statistical Association*, **85**, 328-337 (1990).
31. Stroup, D. F. and Thacker, S. B. 'A Bayesian approach to the detection of aberrations in public health surveillance data', *Epidemiology*, **4**, 435-443 (1993).
32. Stroup, D. F., Wharton, M., Kafadar, K. and Dean, A. G. 'Evaluation of a method for detecting aberrations in public health surveillance data', *American Journal of Epidemiology*, **137**, 373-380 (1993).
33. Wharton, M., Price, W., Hoesly, F., Woolard, D., White, K., Greene, C. and McNabb, S. 'Evaluation of a method for detecting outbreaks of diseases in six states', *American Journal of Preventive Medicine*, **9**, 45-49 (1993).
34. Nobre, F. F. and Stroup, D. F. 'A monitoring system to detect changes in public health surveillance data', *International Journal of Epidemiology*, **23**, 408-418 (1994).
35. Stroup, D. F. 'Special analytic issues', in Teutsch, S. M. and Churchill, R. E. (eds), *Principles and Practice of Public Health Surveillance*, Oxford University Press, New York, 1994.
36. VanBrackle, L. and Williamson, G. D. 'A study of the average run length characteristics of the National Notifiable Diseases Surveillance System', *Statistics in Medicine*, **18**, 3309-3319 (1999).
37. Choi, K. and Thacker, S. B. 'An evaluation of influenza mortality surveillance, 1962-1979. I. Time series forecasts of expected pneumonia and influenza deaths', *American Journal of Epidemiology*, **113**, 215-226 (1981).
38. Helfenstein, U. 'Box-Jenkins modelling of some viral infectious diseases', *Statistics in Medicine*, **5**, 37-47 (1986).
39. Stroup, D. F., Thacker, S. B. and Herndon, J. L. 'Application of multiple time series analysis to the estimation of pneumonia and influenza mortality by age, 1962-1983', *Statistics in Medicine*, **7**, 1045-1059 (1988).
40. Kopjar, B. and Guldvog, B. 'Time variations in injury incidence', *National Institute of Public Health Annals*, **16**, 3-10 (1993).
41. Monteiro, A. B. S., Telles, P. R., Nobre, F. F. and Williamson, G. D. 'A comparison of epidemiological time series forecast methods: Bayesian and ARIMA models' (unpublished manuscript).
42. Frisen, M. 'Evaluations of methods for statistical surveillance', *Statistics in Medicine*, **11**, 1489-1502 (1992).

43. Centers for Disease Control and Prevention. *Summary of Notifiable Diseases, United States 1996, Morbidity and Mortality Weekly Report*, **45**(53), 1997.
44. Montgomery, D. C. *Introduction to Statistical Quality Control*. John Wiley & Sons, New York, 1991.
45. Yourstone, S. A. and Montgomery, D. C. 'A time-series approach to discrete real-time process quality control', *Quality and Reliability Engineering International*, **5**, 309–317 (1989).
46. Montgomery, D. C. and Mastrangelo, C. M. 'Some statistical process control methods for autocorrelated data (with discussion)', *Journal of Quality Technology*, **23**, 179–204 (1991).
47. Makridakis, S., Wheelwright, S. C. and McGee, V. E. *Forecasting: Methods and Applications*. John Wiley & Sons, New York, 1983.
48. Pankratz, A. *Forecasting with Univariate Box–Jenkins Models Concepts and Cases*. John Wiley & Sons, New York, 1983.
49. *SAS/ETS<sup>R</sup> User's Guide*. SAS Institute Inc., Cary, NC, 1993.
50. Shewhart, W. A. *Economic Control of Quality of Manufactured Product*, Van Nostrand, New York, 1931.
51. *SAS/QC<sup>R</sup> Software: Reference*. SAS Institute Inc., Cary, NC, 1990.
52. Roberts, S. W. 'Control chart tests based on geometric moving averages', *Technometrics*, **1**, 239–250 (1959).
53. Hunter, J. S. 'The exponentially weighted moving average', *Journal of Quality Technology*, **18**, 203–210 (1986).
54. Page, E. 'Continuous inspection schemes', *Biometrika*, **41**, 100–115 (1954).
55. Haddad, S. F., Dean, A. G., Williamson, G. D. and Stroup, D. F. *Statistical Software for Public Health Surveillance*, Centers for Disease Control and Prevention, Atlanta, 1994.