Statistical monitoring-based alarming systems in modeling the AIDS epidemic in the United States, 1985-2011

Jingnan Zhang[1,*], Peihua Qiu[1], Xinguang Chen[2]

[1]*Department of Biostatistics, University of Florida, Gainesville, USA;* [2]*Department of Epidemiology, University of Florida, Gainesville, USA*

Abstract: Better decisions for the control of HIV/AIDS and other infectious diseases require better information. The large amount of available public health data makes it possible to extract such information to monitor and predict significant disease events in disease epidemic. The detection of unusual events often involves a combination of a forecasting and a decision mechanism assessing the extent to which an observed event differs significantly from a forecast event. A number of methods and models have been proposed to monitor the trend of infectious disease and to detect unusual events. Although these existing methods and models are useful, many new issues remain to be addressed, including the complicated data structure and the infectious disease dynamics. In this paper, we first reviewed the most commonly used methods and models, including the historical limit method, the time series analysis, the hidden Markov models, and the process control charts. Then, we further discussed issues with the current available methods. To overcome these issues, we introduced the statistical tool using statistical process control, and proposed a new method under that framework. A major feature of the new method is that it prospectively monitors the disease incidence using sequentially collected data over time. It also takes into account a wide variety of longitudinal patterns and possible autocorrelation in the data. We further test this novel method with the recorded data of the number of AIDS cases in different states of US from 1985 to 2011. The results show that our new method is effective in detecting and predicting the time trends of AIDS epidemic for individual states and for US as a whole. Although AIDS data are used in our demonstration, this method can be used for monitoring other infectious diseases.

Keywords: Early Detection; Epidemiology; Incidence Rate; Public Health Surveillance; Sequential Monitoring; Statistical Process Control (SPC); Seasonality.

1. INTRODUCTION

After the first case of the acquired immunodeficiency syndrome (AIDS) in the United States (US) in 1981 [1], this life-threatening epidemic progresses rapidly. By 1993, AIDS had become the leading cause of death among persons 25 to 44 years old and eighth overall among all causes of death [1]. The AIDS epidemic, including outbreaks of new infections and the large number of individuals living with the AIDS virus, create a major public health problem. Trends in the characteristics of AIDS cases are important information to provide to the decision-makers and the general public regarding the epidemic of AIDS. Many methods have been developed and used to detect the trend so that appropriate decisions can be made to prevent and control the epidemic in its early stage. For example, the AIDS case report systems were established in all 50 US states, including the District of Columbia, with the Centers for Disease Control and Prevention (CDC) as the system hub. These reported data provide us the basis for understanding the distribution of the disease over time and by geographic regions. Data from this system are used as a major source to guide public health interventions at the national, state, and local levels. To extract information from this system and other data collection systems to meet the needs for disease prevention and control, public health researchers have developed a wide array of methods to process and utilize the collected data and support the control of AIDS epidemic.

1.1. Surveillance data

Three types of surveillance data are available for disease surveillance and monitoring, including the surveillance and monitoring of HIV/AIDS, and they are (1) number of total cases, (2) incidence rate of a disease, and (3) mortality rate. The number of newly confirmed cases of a disease is recorded periodically (e.g., on a daily, weekly, or monthly basis) through specific surveillance systems in many countries, including the US. It provides the up-to-date information about the number of persons acquiring a specific disease. Such data are useful for disease monitoring in real time.

Sometimes the numbers of new cases alone may not be sufficient to provide a meaningful comparison of disease severity across different places because such data do not take

*Address correspondence to this author at the Department of Biostatistics, University of Florida, Gainesville, USA; Tel/Fax: ++01-612-356-4611; E-mail: zhan1441@ufl.edu

into account the difference of population density at different places. Therefore, the information derived from the total cases cannot tell the difference between a large population with a low disease rate and a small population with a high disease rate. Thus, it is important to relate the number of persons with a disease to the total population when monitoring the disease epidemic in a region. To this end, the incidence rate provides a good alternative, which is defined to be the ratio of the total number of cases diagnosed during a specific time period to the population at risk over the same period.

As to certain health events such as heart disease, cancer, chronic lower respiratory disease, stroke, accident and crime, people are usually interested in the frequency of deaths in a defined population during a specific time interval. In such cases, the mortality rate can be used. In disease monitoring, the mortality rate is defined to be the number of deaths of a specific disease/condition divided by the total population in a related region. Since mortality rates of a population are heavily influenced by the age distribution of the population, age-adjusted mortality rates are often used to ensure the observed differences in deaths over time are not confounded by changes in the age distribution of the population over time [2]. For applications of mortality surveillance, see Aylin et al. [3].

1.2. Statistical surveillance methods

A review of the published studies indicate that among the existing surveillance methods and models, four of them are commonly used in practice to monitor the disease epidemic and estimate the risk of disease outbreak. They are the historical limit method, the time series analysis, the hidden Markov models, and the process control charts, described briefly below.

1.2.1. Historical limit method:

The historical limit method relies on a straightforward comparison of the reported number of health events in the current time period to a summary statistic of past activities, for example, a mean or median. More specifically, it monitors the total number of cases in the current 4-week period and compares it with a baseline value. The baseline value is typically chosen to be the average of the numbers of cases in the preceding, current, and next 4-week periods during the past 5 years. Thus, fifteen values are averaged as an estimate of the baseline value. A ratio is then calculated by dividing the current 4-week total cases by the calculated baseline value [4]. It also assumes that the number of reported cases follows a normal distribution $\mathrm{N}(\mu, \sigma^2)$, and the historical limits of the ratio are

$$1 \pm \frac{2\sigma}{\mu},$$

where the mean μ and the standard deviation σ can be estimated from the fifteen historical incidence values mentioned above. For each disease, if the calculated ratio is greater than the upper historical limit of the ratio, then a potential epidemic alert should be issued and a further epidemiologic investigation should be conducted. Due to its simplicity and interpretability, the historical limits method is quite popular in many health departments. Since 1989, the US Centers for Disease Control and Prevention has applied this method to the disease count data and reported the results in the Morbidity and Mortality Weekly Report [5].

1.2.2. Time series model:

Public health surveillance data are often collected sequentially over time. Thus, these data are usually temporal correlated with seasonal changes. Adaptive methods that model the temporal dynamics have been proposed in the literature to provide forecasts of future incidence rates [6]. Among the adaptive methods, the time series analysis is the most commonly used for disease surveillance.

The well-received autoregressive integrated moving average (ARIMA) method [7] provides a general framework for the time series modeling. For this type of methods to work properly, the times series data must be stationary, i.e., both its mean function and the auto-covariance function are time invariant. When the time series has a non-constant mean, the traditional transformations, such as the time-lag differencing method, are required to generate stationary series. Square root or log transformations are often used when the variance of the time series data changes over time. Lai has implemented several preselected ARIMA models to the 2003 severe acute respiratory syndrome (SARS) epidemic data in Hong Kong [8]. The expected incidence values are estimated based on the one-day forecasts, and the forecasts are then compared with the most recently observed disease incidence value. The ARIMA-based models may not always perform well, particularly in cases when the stationary hypothesis is violated.

Another well-known method is the Serfling's cyclic regression model [9]. This model is used by CDC as the standard algorithm for flu detection and monitoring. Instead of assuming a stationary data, this model uses sine and cosine functions to account for the underlying sinusoidal behavior of the seasonal influenza. By this method, the parameter of the non-epidemic seasonal baseline is first estimated, and then the upper limit of a confidence interval for the sinusoid is used to determine the epidemic threshold in the related time period. Serfling was the person who originally used this model to detect unusual pattern of pneumonia and influenza mortality in 108 US cities.

1.2.3. Hidden Markov models:

Hidden Markov modeling (HMM) approach gains some popularity recently because of their success in detecting the outbreaks of influenza-like illness in some real applications [10,11]. This method characterizes the sequence of surveillance data by assuming that its probability density function depends on the state of an underlying Markov chain. In disease surveillance, the process is assumed to lie into one of two states: an endemic (non-outbreak) state, and an epidemic (outbreak) state. In order to detect the anomaly state of the disease, the most likely sequence of hidden states can be detected by the Viterbi algorithm based on the inference of these two states [12]. In the literature, the two-state Hidden Markov models along with a seasonal trend have been applied to monitor the influenza-like illness data, using a mixture of Gaussian, poliomyelitis counts, and a mixture of Poisson distribution [12]. This method has been demonstrated in examples of several infectious diseases, such as the flu-like disease, Malaria, Leprosy [13], nosocomial infections [14] and Hepatitis A [15].

1.2.4. Statistical process control charts:

The statistical process control (SPC) charts are originated from industrial engineering [16]. They are effective tools for infectious disease monitoring. In the SPC framework, control charts are created and used to monitor a process. The most widely used control charts include (a) Shewhart chart, (b) the cumulative sum (CUSUM) chart, and (c) the exponentially weighted moving average (EWMA) chart. Many biosurveillance systems, such as the Early Aberration Reporting System (EARS), BioSense, ESSENCE and NUCDOHMH, all use the SPC charts to detect disease outbreaks. The SPC charts evaluate the performance of a process sequentially based on all observed data up to the current time point. These methods assume that the observed data are temporally independent and normally distributed. For this reason, it has been shown that SPC methods alone may lead to a relatively poor performance for disease surveillance, if the day-of-the-week effect of a disease is not handled properly [17]. To overcome this limitation, Cowling et al. [18] suggested an upper CUSUM chart using a 7-week buffer interval in the application.

1.3. Challenges to disease surveillance

By using the existing methods described above for disease outbreak detection, there are a number of challenges to deal with the complicated structures of the commonly available HIV/AIDS data for effective detection of a potential event of HIV infection.

First, to detect a HIV/AIDS outbreak, we have to know the baseline pattern of the disease incidence or mortality when there is no disease outbreak. However, proper estimation of the baseline pattern itself is challenging because an observed pattern of HIV/AIDS over time can be affected by many external factors, such as climate change, population mobility, seasonal effects and so on. The regular baseline pattern would not be estimated properly without considering the impact of these time-dependent factors. To develop a robust adaptive monitoring system, we need to consider two different processes of the dynamic changes in HIV/AIDS epidemic: (1) a process that is robust despite random fluctuations in the disease patterns to build the baseline model, and (2) a process that is sensitive to changes that may indicate outbreaks.

Second, the surveillance data for HIV/AIDS are often temporally correlated. If such correlation is ignored in monitoring, false HIV/AIDS outbreaks could be triggered as frequently as every day. Without a careful consideration of the impact of the temporal correlation, any model-based estimates will be subject to increased uncertainty and information bias, both of which will jeopardize our effort to monitor the HIV/AIDS epidemic. The autoregressive integrated moving average (ARIMA) method is recommended to deal with the temporal correlation [19,20]; however, this approach is difficult, if not impossible to implement in an automated way in practice because of the non-stationary nature of the surveillance data. The mean and variance structures of the surveillance data tend to change over time, violating the stationary hypothesis of the ARIMA approach.

Another key task when constructing a surveillance system is to make use of all available data, including the new data and the historical data, so that new information can be used to clarify situational awareness of public health monitors [21]. To this end, it is critical that the designed system can offer the best potential for early intervention and prevention while the false alarm rate is controlled at a low level. Our literature review indicates that most surveillance methods use the current data to make inference about the epidemic or outbreak of a disease without considering the historical data. In such cases, an outbreak in HIV/AIDS will not be detected if the disease incidence or motality showed a gradual increasing trend at the beginning. Ideally, all the historical data should be used for accumulating the evidence of an irregular pattern. With this approach, any outbreak signals can be triggered more efficiently without much delay. The aforementioned SPC methods, including the cumulative sum (CUSUM) and the exponentially weighted moving average (EWMA) charts, provide an effective tool to deal with this issue. The control charts evaluate the performance of a process sequentially based on all observed data up to the current time. But, to use these charts, it is conventionally assumed that the process distribution is unchanged over time when the process is stable (i.e., the process is in-control (IC)). This is obviously violated in HIV/AIDS applications because their incidence rates would change over time even in cases with no disease outbreaks. Therefore, the traditional SPC charts need to be modified properly before they can be applied to the current problem.

1.4. Purpose of this study

In this paper, we report our work on the development and verification of the adaptive monitoring-based alarming system for HIV/AIDS monitoring. The method we reported here can be considered as a modification of the dynamic screening system (DySS) originally proposed by Qiu and Xiang [22]. In this new method we attempted to address the aforementioned three challenges (i.e., addressing both the baseline pattern and online monitoring, dealing with temporal auto-correlation, and inclusion of both the current and history data). The ultimate goal is to develop a new tool for AIDS epidemic surveillance and control. To illustrate the use of our proposed monitoring scheme, the US AIDS diagnostic data from 1985 to 2011 are used and the related results are presented.

2. AN ADAPTIVE MONITORING-BASED SYSTEM

We propose an adaptive monitoring-based alarming system consisting of the following three steps:

(i) *Detrend* – The baseline pattern is first estimated by a nonparametric longitudinal model, and the estimated baseline pattern is then eliminated from the observed data,

(ii) *Decorrelation* – The temporal autocorrelation in the detrended data is modeled by an ARIMA model, and the estimated autocorrelation is eliminated from the detrended data, and

(iii) *Sequential Monitoring* – The adjusted data obtained in step (ii) are then sequentially monitored by a SPC chart.

The main focus of step (i) is to estimate the baseline longitudinal pattern of a disease in question and remove it from the observed data before monitoring. Such baseline longitudinal pattern can usually be explained by the seasonality and other factors that are not our major interest [23, 24]. The baseline pattern will be estimated by the nonparametric regression techniques. After the estimated baseline pattern is removed from the observed data, the mean and variance of the detrended data should remain stable under the non-epidemic condition. So, the underlying assumption of stationarity in the time series analysis in step (ii) is valid. The ARIMA model used in step (ii) is mainly for removing autocorrelation in the observed data. After this step, the detrended and decorrelated data should be independent at different time points and stationary if there are no disease outbreaks. A conventional SPC chart can then be applied for online monitoring of the disease incidence. Details about our three-step strategy are described in the following several parts.

2.1. Detrend

In order to use the proposed monitoring scheme, the baseline longitudinal pattern of the surveillance data over time needs to be estimated properly. To this end, assume that the observed disease incidence rate y follows the nonparametric regression model

$$y(t_{ij}) = \mu(t_{ij}) + \varepsilon(t_{ij}), \qquad \text{for } i = 1, 2, \dots, m, \qquad (1)$$
$$j = 1, 2, \dots, n,$$

where $t_{ij} \in [0, T]$ is the jth observation time at location i, $\mu(t_{ij})$ is the mean of $y(t_{ij})$, and $\varepsilon(t_{ij})$ is the error term. We further assume that the error term is the sum of two independent components, i.e, $\varepsilon(t_{ij}) = \varepsilon_0(t_{ij}) + \varepsilon_1(t_{ij})$, where $\varepsilon_0(\cdot)$ is a random process with mean 0 and covariance function $V_0(s,t), s, t \in [0,T]$. And, $V_0(s,s)$ is denoted as $\sigma_0^2(s)$. In this decomposition, $\varepsilon_1(\cdot)$ denotes the pure measurement error with mean 0 and variance $\sigma_1^2(s)$, and $\varepsilon_0(\cdot)$ denotes all possible covariates that may affect y but are not included in model (1). The estimator of $\mu(t_{ij})$ is proposed by Chen et al. [25] and Pan et al. [26] based on the local pth-order polynomial kernel smoothing procedure. Similarly, the variance functions $\sigma_0^2(s)$ and $\sigma_1^2(s)$ can be estimated. By modeling the error term in this way, it allows for temporal correlation among observed data within the area without specifying the autocorrelation structure. This model is flexible enough to accommodate a wide range of correlation in the data, and adjust for the baseline pattern that can vary from site to site. After model (1) is estimated, the observed data at each location are adjusted by first subtracting the estimated mean and then dividing the estimated standard deviation of $y(t_{ij})$. The resulting standardized data are denoted as $r(t_{ij})$. Note that through the above data standardization, the heterogeneity problem in the mean and variance functions over time and at different places has been addressed.

2.2. Decorrelation

Before start modeling the temporal correlation in the standardized data, we suggest performing Ljung-Box test [27] to assess whether a strong temporal correlation exists. By this test, a small p-value indicates a strong evidence of dependence among data. In that case, the temporal correlation among the standardized data can be described by an ARIMA model. We then can use the Box-Jenkins technique to develop a forecasting model [28]. The Box-Jenkins approach to time series forecasting requires an adequate stochastic ARIMA model of the following form to describe the time series at location i:

$$\Phi_p(B)\nabla^d r(t_{ij}) = \Theta_q(B)\epsilon(t_{ij}), \qquad j = 1, 2, \dots, n, \qquad (2)$$

where $\Phi_p(B) = (1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p)$ is an autoregressive polynomial of order p, ∇ is the backward difference operator, d is the order of the first difference, $\Theta_q(B) = (1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q)$ is a moving average polynomial of order q, B is the backshift operator, and $\epsilon(t_{ij})$ is a sequence of independent and identically distributed (i.i.d.) random errors with mean 0 and unknown variance. To develop and select an appropriate ARIMA model for each series, we suggest choosing d using the successive KPSS unit-toot test [29]. The KPSS test for testing the null hypothesis that the time series model is stationary is performed for integer $d \geq 1$, and we choose the smallest integer value of d such that the corresponding p-value of the test is less than a prespecified significance level α. Throughout this paper, α is set to be 0.05. Once d is determined, p and q are chosen by minimizing the AICc criterion [29]. Note that the ARIMA model for each series is chosen in cases when no outbreak is present. Given the most recent observations, the one-step-ahead forecast errors based on the fitted chosen model can then be used for sequential monitoring.

2.3. Sequential monitoring

In this part, we discuss the on-line monitoring of the forecast errors for detecting any upward shift in the disease incidence rate as quickly and accurately as possible. To this end, SPC charts have been frequently used. Recently, such charting methods have been demonstrated to be useful in public-health-related applications as well [30]. Among these commonly used control charts, the CUSUM chart has been proved to have certain good theoretical properties [31]. It is optimal in the sense that it detects the mean shift of a pre-specified size with the least delay on average if the observations are independent and normally distributed. To measure the performance of a control chart, the average run length (ARL), defined as the average number of time points before an alarm is given, is often used. The control chart is usually presented by a plot of some function of the process measurements, called the charting statistic, against time. The computed value of the charting statistic is then compared to a pre-specified control limit. At the first time when the charting statistic exceeds the control limit, the chart gives an alarm, indicating that some systematic shift may have occurred and some actions should be taken to fix the problem. The design of a control chart (i.e., selection of the control limit and other parameters) is a compromise between the risks of missing real shifts and giving false alarms. When the disease process is regular (or called *in-control* (IC) in the SPC literature), all alarms are false alarms. The distribution of the false alarms is often summarized by the IC ARL, denoted as ARL_0. On the other hand, when the disease process is in the epidemic state (or called *out-of-control* (OC)), the OC ARL, denoted as ARL_1, measures the excessive delay in detecting a true shift. The ARL_0 and ARL_1 are well-defined measures that can be computed easily using existing software packages. They are the standard metrics for measuring the performance of the CUSUM chart. Intuitively, for an IC process, the ARL_0 value should be as large as possible. For example, if the observations from an IC process are independent and have a standard normal distribution and $\text{ARL}_0 = 370$, it means that even if the process remains IC, on average an observation exceeding the control limit and triggering a signal every 370 observations. On the other hand, if the process is OC, ARL_1 should be as small as possible, meaning that the control chart signals the shift as soon as possible after the shift occurs. Unfortunately, both metrics cannot be small at the same time. For instance, in cases when the ARL_0 value is large, the ARL_1 value would be relatively large as well, and vice versa. To make a trade-off, the ARL_0 value is usually fixed at a given level (e.g., 200), and then we try to make the ARL_1 value as small as possible.

In this paper, we mainly use the CUSUM chart because of its good theoretical properties discussed above. To detect an upward mean shift for any sequence $\{\mathrm{e_j}, j = 1, 2, 3, \ldots\}$, the charting statistic of the upward CUSUM chart is defined by

$$C_j^+ = \max\left(0, C_{j-1}^+ + e_j - k\right), \qquad \text{for } j \geq 1, \qquad (3)$$

where $C_0^+ = 0$, and $k > 0$ is called the allowance constant. The chart gives a signal of an upward mean shift if

$$C_j^+ > h, \qquad (4)$$

where $h > 0$ is called the control limit. Similarly, to detect a downward mean shift, the charting statistic is defined by

$$C_j^- = \min\left(0, C_{j-1}^- + e_j + k\right), \qquad \text{for } j \geq 1, \qquad (5)$$

where $C_0^- = 0$. The chart gives a signal of a downward mean shift if

$$C_j^- < -h, \qquad (6)$$

We can also combine the two one-sided control charts for detecting both upward and downward shifts. Usually, the allowance k is specified beforehand. It has been well demonstrated in the literature that large k values are effective for detecting large shifts and small k values are effective for detecting small shifts. The control limit h is chosen such that a pre-specified ARL_0 value is reached. Then, the chart performs better for detecting a given shift if its ARL_1 value is smaller. ARL_1 usually depends on the size of the shift, denoted as δ. It has been proved in the literature [31] that the best k value is $\delta/2$ if process observations are independent and normally distributed. Since the data obtained from step 2 are approximately stationary and independent, the assumption of normality is the only main fact that we should consider. When the normality assumption is not valid but a set of IC dataset is available, we suggest using a bootstrap (or resampling) procedure to search for the control limit of the CUSUM charts [32, 33].

3. APPLICATION TO AIDS SURVEILLANCE IN THE UNITED STATES

Here we illustrate the use of our monitoring scheme for monitoring the incidence rate of AIDS in US. The numbers of reported AIDS cases are geographically aggregated to the 50 states and District of Columbia each year during 1985 – 2011. As the denominator for calculating the incidence rate, we use the official population estimates for each state obtained from the official website of the U.S. Census Bureau. Figure (1) shows the observed incidence rate of AIDS for each of the 50 states and District of Columbia over the 27-year follow-up period. From the figure, it can be seen that the incident rates increase from 1985 to 1993, and a peak is observed around 1993 for many states. We hope that their increasing trends can be detected early by our proposed monitoring scheme so that some medical interventions can be applied in a timely manner.

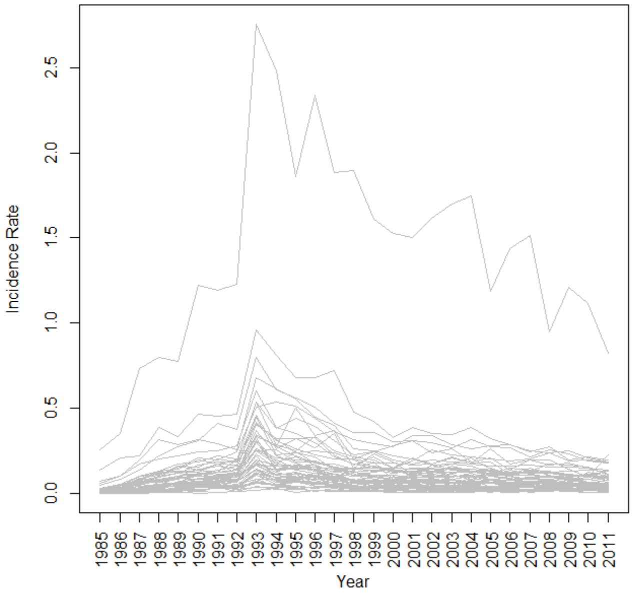

Fig. (1). The incidence rate of AIDS per 100,000 people for each state in the United States during 1985 – 2011.

To estimate the baseline model (1), the local linear kernel smoothing procedure [34, Chap 2] is applied to the IC data obtained in the following way. For each state, the 25th and 75th percentiles of the AIDS incidence rate are first computed. Then, these percentiles are averaged among all 51 states, respectively. The resulting central 50% confidence interval for the AIDS incidence rate is [0.0045, 0.3880]. Then, for each state, all observed incidence rates that are within this interval are regarded as IC. From the defined IC data, an estimated mean function of the AIDS incidence rate is obtained as described in step 1. Once a time series model is estimated from the detrended IC data, it is applied to all the detrended data for eliminating the autocorrelation. After the observed AIDS incidence data are standardized (or, detrended) and decorrelated, the CUSUM chart can be applied to the resulting forecasting errors. In this case, we noted that the data do not seem to follow a normal distribution. In such cases, the commonly used control limit h computed for normal data cannot be used, and it needs to be numerically searched by an algorithm. We use a bootstrap resampling algorithm to determine the value of h such that the pre-specified ARL_0 is reached. More specifically, we resample with replacement the IC detrended and decorrelated data and numerically search for the appropriate value of h. After the value of h is computed, the conventional CUSUM chart can be readily applied. In this paper, k is set to be 0.5, and ARL_0 is fixed at 200.

4. RESULTS

First, the IC data in the 51 states (light grey lines) and the estimated baseline function (black dashed line) are presented in Figure (2). From the plot, it can be seen that the estimated baseline function captures the general longitudinal trends of the AIDS incidence rates well. We monitor the detrended and decorrelated data for each state by the two-sided CUSUM chart and 14 out of 51 states triggered the outbreak signals during 1985 – 2011. The signal times are listed in Table 1 and shown in Figure (3). From the table and the figure, it can be seen that the District of Columbia is the first state in getting the signal in 1986. Its numbers of AIDS cases are consistently at high levels considering its relatively small population. Its AIDS incidence rate is more than 9 times the national average rate. The CUSUM chart gives signals during 1987 – 1989 for the states New York, New Jersey, Florida, and California. This result agrees with the analysis done by CDC which reveals that new AIDS diagnoses are concentrated primarily in metropolitan areas (81% in 2011), with New York, Los Angeles, and Miami topping the rank. Several states get signals around 1993, several others get signals during 1996 – 1999, and no states get signals after 1999. Figure (4) shows the charting statistic C_j^+ in (3) and the observed AIDS incidence rate for the state of California. The left panel shows the charting statistic in (3), with the dashed line denoting the numerically searched control limit h. It can be seen that the charting statistic rises and then stabilizes, implying that the AIDS incidence rate increases quite quickly first and then stabilizes or decreases. The right panel plots the observed AIDS incidence rate for California during 1985-2011. The vertical line indicates the signal time. It can be seen that the signal is given way before 1993 when an AIDS outbreak is quite obvious. Therefore, the CUSUM chart is quite effective for early detection of the AIDS outbreaks.

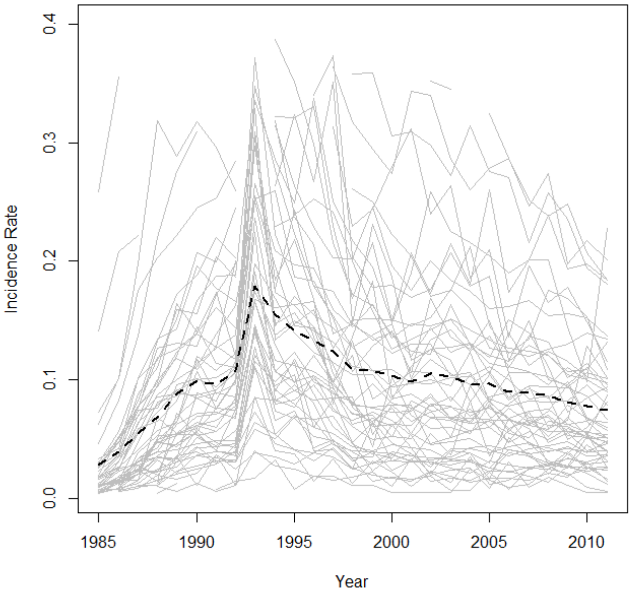

Fig. (2). The in-control (IC) AIDS incidence rates and their estimated baseline function. The grey solid lines are the IC incidence rates of each state, and the black dotted line is the estimated baseline function.

Table 1. Signal times of the CUSUM chart for certain states in the U.S. during 1985 – 2011.

State	Signal time of outbreak	State	Signal time of outbreak
District of Columbia	1986	Georgia	1993
New York	1987	Nevada	1993
New Jersey	1988	Texas	1993
Florida	1989	Delaware	1994
California	1989	Louisiana	1996
Maryland	1993	South Carolina	1997
Connecticut	1993	Massachusetts	1999

5. DISCUSSION AND CONCLUSION

We have described a three-step procedure for early detection of the AIDS outbreak in the United State. This procedure can accommodate the baseline longitudinal trend of the disease incidence rate and the temporal autocorrelation in the observed data as well. Our numerical results show that it is effective in practice. In our future research, we will generalize this approach in several different directions. First, the AIDS incident rates are often influenced by certain environmental variables, such as medical conditions, educational levels, social-economic status, and so forth. Such AIDS-related covariates will be incorporated in model (1) when we estimate the baseline pattern. Second, in the current procedure, the spatial correlation in the observed data has not been taken care of yet. We will develop effective methods for sequentially monitoring the spatio-temporal patterns of the AIDS incidence rate and for more effective detection of the disease outbreaks.

CONFLICT OF INTEREST

None of the authors have conflicts of interest to declare.

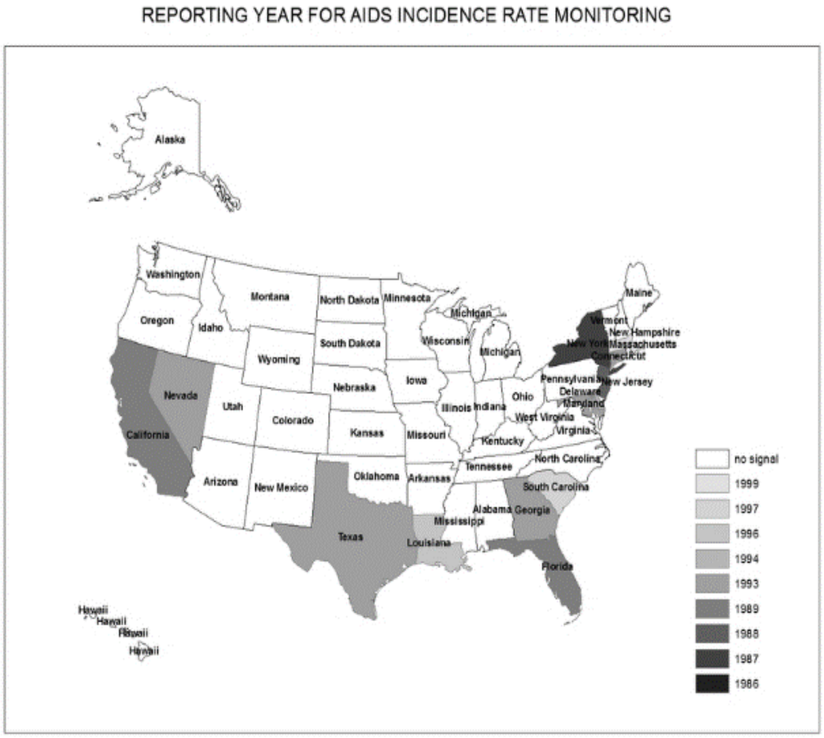

Fig. (3). Signal times of different states are shown by different colors: the darker, the earlier.

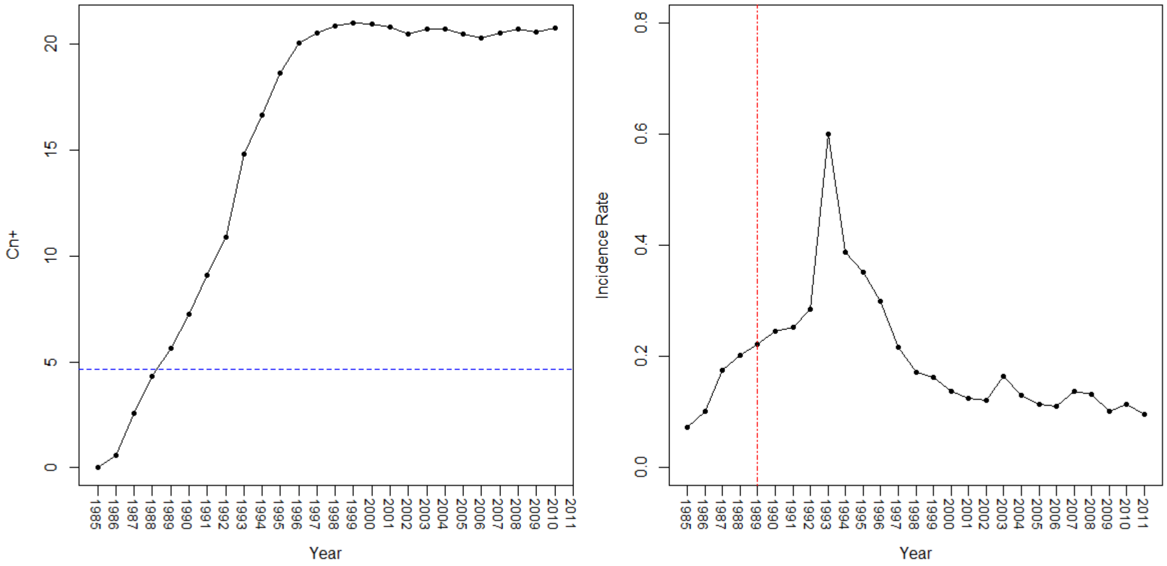

Fig. (4). The charting statistics $C_j^+$and the AIDS incidence rate for the state of California. Left panel: charting statistic in (3) is plotted, and the dashed line is the numerically searched control limit h, in cases when ARL$_0$ is fixed at 200 and k is set to be 0.5. Right panel: the observed AIDS incidence rate of California during 1985 – 2011, and the vertical line indicates the signal time.

REFERENCES

[1] Centers for Disease Control and Prevention, Prevention. HIV and AIDS--United States, 1981-2000. MMWR Morbidity and mortality weekly report 2001; 50(21):430-4.

[2] Notestein FW. Mortality, fertility, the size-age distribution, and the growth rate. Demographic and economic change in developed countries: Columbia University Press 1960; pp. 261-84.

[3] Aylin P, Best N, Bottle A, Marshall C. Following Shipman: a pilot system for monitoring mortality rates in primary care. Lancet 2003; 362(9382):485-91.

[4] Stroup DF, Williamson GD, Herndon JL, Karon JM. Detection of aberrations in the occurrence of notifiable diseases surveillance data. Statistics in medicine 1989; 8(3):323-9; discussion 31-2.

[5] Centers for Disease Control and Prevention. Proposed changes in format for presentation of notifiable disease report data. MMWR Morbidity and mortality weekly report 1989; 38(47):805-9.

[6] Lawson AB, Kleinman K. Spatial and Syndromic Surveillance for Public Health. Wiley 2005.

[7] Box GEP, Jenkins GM. Time series analysis: forecasting and control: Holden-Day 1976.

[8] Lai D. Monitoring the SARS epidemic in China: a time series analysis. Journal of Data Science 2005; 3(3):279-93.

[9] Serfling RE. Methods for current statistical analysis of excess pneumonia-influenza deaths. Public health reports 1963; 78(6):494.

[10] Rath TM, Carreras M, Sebastiani P. Automated detection of influenza epidemics with hidden Markov models. Advances in Intelligent data analysis V: Springer 2003; pp. 521-32.

[11] Sebastiani P, Mandl KD, Szolovits P, Kohane IS, Ramoni MF. A Bayesian dynamic model for influenza surveillance. Statistics in medicine 2006; 25(11):1803-16.

[12] Le Strat Y, Carrat F. Monitoring epidemiologic surveillance data using hidden Markov models. Statistics in medicine 1999; 18(24):3463-78.

[13] Jamshidi Orak R, Mohammad K, Pasha E, Sun W, Nori Jalyani K, Rasolinejad M, et al. Modeling the spread of infectious diseases based the Bayesian approach. Journal of School of Public Health and Institute of Public Health Research 2007; 5(1):7-15.

[14] Cooper B, Lipsitch M. The analysis of hospital infection data using hidden Markov models. Biostatistics 2004; 5(2):223-37.

[15] Watkins RE, Eagleson S, Veenendaal B, Wright G, Plant AJ. Disease surveillance using a hidden Markov model. BMC medical informatics and decision making 2009; 9(1):39.

[16] Qiu P. Introduction to Statistical Process Control. 1st ed. CRC Press 2013.

[17] Jackson ML, Baer A, Painter I, Duchin J. A simulation study comparing aberration detection algorithms for syndromic surveillance. BMC Medical Informatics and Decision Making 2007; 7(1):6.

[18] Cowling BJ, Wong IO, Ho L-M, Riley S, Leung GM. Methods for monitoring influenza surveillance data. International Journal of Epidemiology 2006; 35(5):1314-21.

[19] Yu H-K, Kim N-Y, Kim SS, Chu C, Kee M-K. Forecasting the number of human immunodeficiency virus infections in the Korean population using the autoregressive integrated moving average model. Osong public health and research perspectives 2013; 4(6):358-62.

[20] Aboagye-Sarfo12 P, Cross J, Mueller U. Application of Intervention Analysis to Incidence Cases of HIV Infection in Ghana. Available at: https://www.statssa.gov.za/isi2009/ScientificProgramme/IPMS/0745.pdf. Accessed February 22, 2015.

[21] Shmueli G. To explain or to predict? Statistical science 2010; 25(3):289-310.

[22] Qiu P, Xiang D. Univariate Dynamic Screening System: An Approach For Identifying Individuals With Irregular Longitudinal Behavior. Technometrics 2014; 56(2):248-60.

[23] McQuillan GM, Khare M, Karon JM, Schable CA, Vlahov D. Update on the seroepidemiology of human immunodeficiency virus in the United States household population: NHANES III, 1988-1994. JAIDS Journal of Acquired Immune Deficiency Syndromes 1997; 14(4):355-7.

[24] Bacchetti P, Segal MR, Jewell NP. Backcalculation of HIV infection rates. Statistical Science 1993; 8(2):82-101.

[25] Chen K, Jin Z. Local polynomial regression analysis of clustered data. Biometrika 2005; 92(1):59-74.

[26] Pan J, Ye H, Li R. Nonparametric regression of covariance structures in longitudinal studies. Technical Report: School of Mathematics, University of Manchester, UK 2009.

[27] Shumway RH, Stoffer DS. Time series analysis and its applications: with R examples: Springer Science & Business Media 2010.

[28] Box GE, Jenkins GM. Time series analysis: forecasting and control, revised ed: Holden-Day 1976.

[29] Kwiatkowski D, Phillips PC, Schmidt P, Shin Y. Testing the null hypothesis of stationarity against the

alternative of a unit root: How sure are we that economic time series have a unit root? Journal of econometrics 1992; 54(1):159-78.

[30] Woodall WH. The use of control charts in health-care and public-health surveillance. Journal of Quality Technology 2006; 38(2):89-104.

[31] Moustakides GV. Optimal stopping times for detecting changes in distributions. The Annals of Statistics 1986; 14(4):1379-87.

[32] Edopka I, Ogbeide E. Bootstrap approach control limit for statistical quality control. International Journal of Engineering Science Invention 2013; 2(4): 28-33.

[33] Chatterjee S, Qiu P. Distribution-free cumulative sum control charts using bootstrap-based control limits. The Annals of Applied Statistics 2009; 3(1):349-369.

[34] Qiu P. Image processing and jump regression analysis. 1st ed. John Wiley & Sons 2005.