

CUSUM Residual Charts for Monitoring Enterovirus Infections

Huifen Chen and Yu Chen

Abstract We consider the syndromic surveillance problem for enterovirus (EV) like cases. The data used in this study are the daily counts of EV-like cases sampled from the National Health Insurance Research Database in Taiwan. To apply the CUSUM procedure for syndromic surveillance, a regression model with time-series error-term is used. Our results show that the CUSUM chart is helpful to detect abnormal increases of the visit frequency.

Keywords CUSUM chart · Enterovirus syndrome · Regression analysis · Syndromic surveillance

1 Introduction

The two major epidemic peaks for enterovirus (EV) diseases in Taiwan occur in May to June and September to October yearly according to the historical statistics from Centers for Disease Control in Taiwan (Taiwan CDC). In 1998 the EV infection caused 78 deaths and 405 severe cases in Taiwan (Ho et al. 1999). Early detection of outbreaks is important for timely public health response to reduce morbidity and mortality. By early detecting the aberration of diseases, sanitarians can study or research into the causes of diseases as soon as possible and prevent the cost of the society and medical treatments. Traditional disease-reporting surveillance mechanisms might not detect outbreaks in their early stages because laboratory tests usually take long time to confirm diagnoses.

Syndromic surveillance was developed and used to detect the aberration of diseases early (Henning 2004). The syndromic surveillance mechanism is to

H. Chen (✉) · Y. Chen
Department of Industrial and Systems Engineering, Chung-Yuan University, Chungli,
Taoyuan, Taiwan
e-mail: huifen@cycu.edu.tw

collect the baseline data of prodromal phase symptoms and detect the aberration of diseases from the expected baseline by placing the variability of data from the expected baseline. Such surveillance methods include the SPC (statistical process control) based surveillance methods, scan methods and forecast-based surveillance methods (Tsui et al. 2008). See Sect. 2 for literature review.

In this work, we apply the CUSUM residual chart for detecting the abnormal increases of EV-like cases in Taiwan. Since the daily visits of the EV-like syndrome are time series data with seasonal effect, we use a regression model with an time-series error term to model the daily counts from ambulatory care clinic data. The residuals are then used for the CUSUM chart to detect unusual increase in daily visits. The test data are the 2003–2006 ambulatory care clinic data from the National Health Insurance Research Database (NHIRD) in Taiwan.

This paper is organized as follows. In Sect. 2, we review related literature. In Sect. 3, we summarize the data, propose a regression model whose error term follows an ARIMA model, and construct the CUSUM chart using the residuals. The conclusions are given in Sect. 4.

2 Literature Review

We review here the syndromic surveillance methods including the forecast-based, scan statistics, and SPC-based methods.

The forecast-based methods are useful to model non-stationary baseline data before monitoring methods can be applied. Two popular forecasting methods are time-series and regression models. Goldenberg et al. (2002) used the AR (Auto Regressive) model to forecast the over-the-counter medication sales of the anthrax and built the upper prediction interval to detect the outbreak. Reis and Mandl (2003) developed generalized models for expected emergence-department visit rates by fitting historical data with trimmed-mean seasonal models and then fitting the residuals with ARIMA models. Lai (2005) used three time series models (AR, a combination of growth curve fitting and ARMA error, and ARIMA) to detect the outbreak of the SARS in China.

Some works fitted the baseline data with a regression model first and then fitted the residuals with a time-series model because the baseline data may be affected by the day of the week and/or holiday factors. Miller et al. (2004) used the regression model with AR error to fit the influenzalike illness data in an ambulatory care network. The regression terms include weekend, holiday and seasonal adjustments (sine and cosine functions). Therefore, they used the standardized CUSUM chart of the residuals for detecting the outbreak. Fricker et al. (2008) applied the adaptive regression model with day-of-the-week effects using an 8-week sliding baseline and used the CUSUM chart of the adaptive regression residuals to compare with the Early Aberration Reporting System (EARS). They showed that the CUSUM chart applied to the residuals of adaptive regressions performs better than the EARS method for baseline data with day-of-the-week effects.

The scan statistics method is widely used in detecting the clustering of diseases. Scan statistics methods can be used in temporal, spatial and spatiotemporal surveillance. Heffernan et al. (2004) applied the scan statistic method to monitor respiratory, fever diarrhea and vomiting syndromes by the chief complaint data of the emergency department. They used this method in the citywide temporal and the spatial clustering surveillances. Han et al. (2010) compared CUSUM, EWMA and scan statistics for surveillance data following Poisson distributions. The results showed that CUSUM and EWMA charts outperformed the scan statistic method.

Recently the control charts have been applied in health-care and public-health surveillance (Woodall 2006). The SPC methods were first applied in the industrial statistical control (Montgomery 2005). Since the Shewhart chart is insensitive at detecting small shifts, CUSUM and exponentially weighted moving average (EWMA) charts are more commonly used in public health surveillance than the Shewhart chart. Hutwagner et al. (1997) developed a computer algorithm based the CUSUM chart to detect salmonella outbreaks by using the laboratory-based data. Morton et al. (2001) applied Shewhart, CUSUM and EWMA charts to detect and monitor the hospital-acquired infections. The result shows that Shewhart and EWMA work well for bacteremia and multiresistant organism rates surveillance and that CUSUM and Shewhart charts are suitable for monitoring surgical infection. Rogerson and Yamada (2004) applied a Poisson CUSUM chart to detect the lower respiratory tract infections for 287 census tracts simultaneously. Cowling et al. (2006) adopted the CUSUM chart with 7-week buffer interval for monitoring influenza data form Hong Kong and the United States and compared with time series and regression models. Woodall et al. (2008) show that the CUSUM chart approach is superior to the scan statistics.

3 Methods

3.1 Data Source

The data used in this study are the 2003–2006 daily counts (i.e. the number of daily visits) of EV-like cases for 160,000 people sampled from the National Health Insurance Research Database (NHIRD) by the Bureau of National Health Insurance, Taiwan. Patients' diagnoses in NHIRD were encoded using the ICD-9-CM (International Classification of Diseases, 9th Revision, Clinical Modification Reference) code. In this study, the ICD-9 codes of the EV-like syndrome are adopted from Wu et al. (2008) as listed in Appendix A.

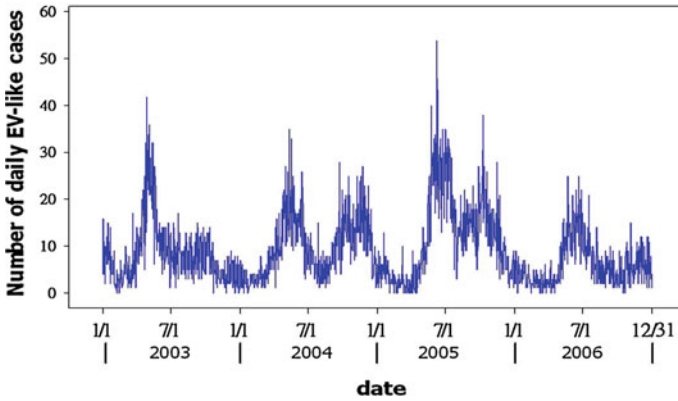


Fig. 1 The daily counts of the EV-like cases from 2003 to 2006

3.2 Data Summary

Here we summarize the daily counts of EV-like cases from 2003 to 2006 with population size 160,000. Figure 1, the run chart of the daily counts, shows that the daily counts are time-series data with seasonal variation. In general, the major epidemic peak occurs in May and June and a smaller peak occurs in September and October. Among the four years, the epidemic peaks are highest in 2005 and lowest in 2006. The day-of-the-week effect also exists. For the age effect, since more than 80 % of the EV-like cases are children younger than 6 years old, we do not consider the age effect in this study.

3.3 CUSUM Charts

Since the daily counts are time series data with seasonal variation, we use the regression model with an ARIMA error term to fit the daily counts of the EV-like cases. For normality, we first use the Box-Cox transformation to transform the daily counts data. The predictor variables are set based on the day-of-the-week, month-of-the-year, and trend effects.

The residuals calculated from the fitted regression model with an ARIMA error term can be used to construct an upper one-sided standardized CUSUM chart (Montgomery 2005) for detecting abnormal increases in daily counts of EV-like cases. Like Miller et al. (2004), we set the control limits so that the in-control average run length is 50.

To illustrate the surveillance method, we use the 2003 and 2004 daily counts data to fit a regression model and then use the 2005 data to construct CUSUM charts and forecasted values.

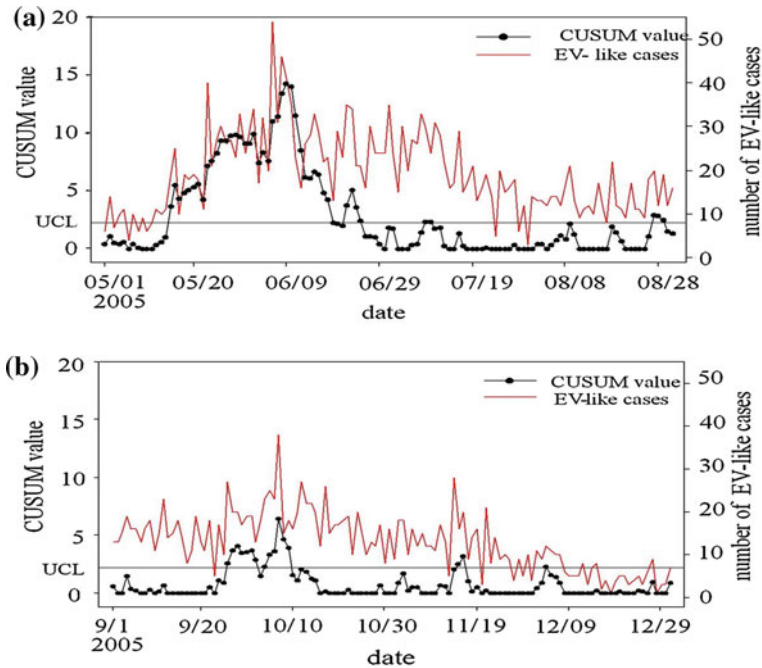


Fig. 2 The CUSUM chart for (a) May 1 to August 31 and (b) Sept. 1 to Dec. 31 in 2005

Figure 2 contain the upper one-sided standardized CUSUM charts for two periods—May 1 to August 31 in Subfigure (a) and September 1 to December 31 in Subfigure(b)—containing epidemic peaks in 2005. The number of EV-like cases and upper control limit (UCL) of the CUSUM chart are also shown. Figure 2 shows that the epidemic outbreak that occurred in May 2005 is detected quickly by the CUSUM chart. The smaller epidemic outbreak occurring at the end of September 2005 is also detected.

Using the fitted regression model, we can construct the 1-steps-ahead forecast value. Figure 3 compares the 2005 actual daily numbers of EV-like cases with its forecasted values for the period (May 1 to December 31) including high seasons. The x-axis is the date and y-axis is the actual (black line)/forecasts (gray line) number of daily EV-like cases. Figure 3 shows that the difference is higher during the peak of infection than in the low season.

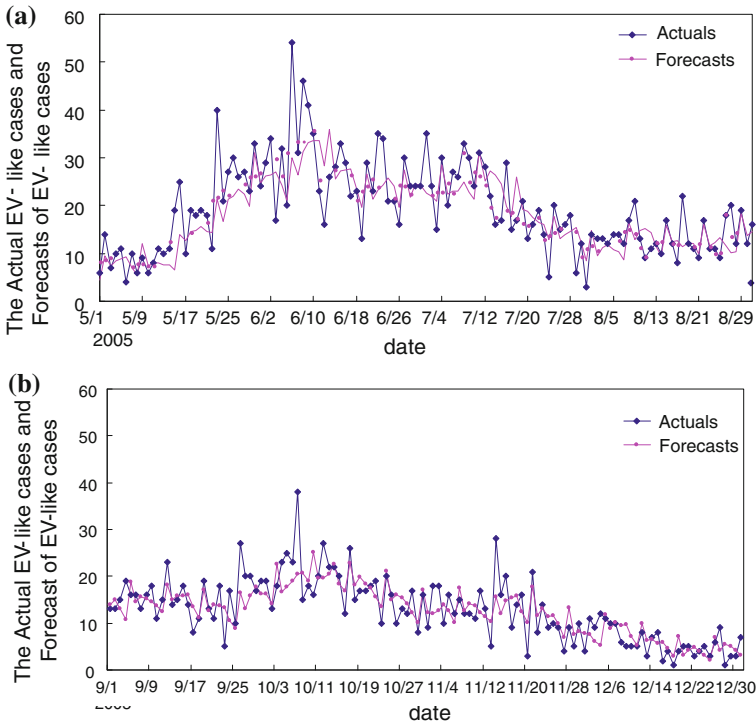


Fig. 3 The actual daily EV-like cases and its forecasts for year 2005: May 1 to Aug. 31 in Subfigure (a) and Sept. 1 to Dec. 31 in Subfigure (b)

4 Conclusions

This paper discusses the implementation of CUSUM residual charts for monitoring daily counts of EV-like cases. The population size is 160,000. Before using the CUSUM chart, we fit a regression model with an ARIMA error term to the daily counts data. The numerical results indicate that the CUSUM residual chart seems to work well in showing unusual increases in daily counts of EV-like cases.

Our fitted regression model is based on historical data of the past two years. The time window can be longer so that more data can be used for model fitting. The shortage though is that the coefficient estimates would have larger variance and hence the prediction interval would be wider. Furthermore, the behavior of daily counts may not be the same each year, using historical data occurring long ago may hurt the prediction accuracy for the future observations.

Acknowledgments This study is based in part on data from the National Health Insurance Research Database provided by the Bureau of National Health Insurance, Department of Health and managed by National Health Research Institutes. The interpretation and conclusions contained herein do not represent those of Bureau of National Health Insurance, Department of Health or National Health Research Institutes in Taiwan.

Appendix A: The ICD-9-CM code of EV-like syndrome

In this study, we adopt the EV-like syndrome definitions from Wu et al. (2008). The ICD-9 codes are listed below.

ICD9	Description
074	Specific diseases due to Coxsackie virus
079.2	Coxsackie virus: infection NOS
047.0	Coxsackie virus: meningitis
074.0	Herpangina Vesicular pharyngitis
074.1	Epidemic pleurodynia, Bornholm disease, Devil's grip; Epidemic: myalgia, myositis
074.2	Coxsackie carditis
074.20	Coxsackie carditis unspecified
074.21	Coxsackie pericarditis
074.22	Coxsackie endocarditis
074.23	Coxsackie myocarditis, Aseptic myocarditis of newborn
074.3	Hand, foot, and mouth disease Vesicular stomatitis and exanthem
074.8	Other specified diseases due to Coxsackie virus, Acute lymphonodular pharyngitis

References

- Cowling BJ, Wong IOL, Ho L-M, Riley S, Leung GM (2006) Methods for monitoring influenza surveillance data. *Int J Epidemiol* 35:1314–1321
- Fricker RD Jr, Hegler BL, Dunfee DA (2008) Comparing syndromic surveillance detection methods: EARS' versus a CUSUM-based methodology. *Stat Med* 27:3407–3429
- Goldenberg A, Shmueli G, Caruana RA, Fienberg SE (2002) Early statistical detection of anthrax outbreaks by tracking over-the-counter medication sales. *Proc Natl Acad Sci USA* 99:5237–5240
- Han SW, Tsui K-L, Ariyajunya B, Kim SB (2010) A comparison of CUSUM, EWMA, and temporal scan statistics for detection of increases in Poisson rates. *Qual Reliab Eng Int* 26:279–289
- Heffernan R, Mostashari F, Das D, Karpati A, Kuldorff M, Weiss D (2004) Syndromic surveillance in public health practice, New York City. *Emerg Infect Dis* 10:858–864
- Henning KJ (2004) What is syndromic surveillance? *Morb Mortality Wkly Rep* 53(Supplement):7–11
- Ho M, Chen ER, Hsu KH (1999) An epidemic of enterovirus 71 infection in Taiwan. *Medicine* 341:929–935
- Hutwagner LC, Maloney EK, Bean NH, Slutsker L, Martin SM (1997) Using laboratory-based surveillance data for prevention: an algorithm for detecting Salmonella outbreaks. *Emerg Infect Dis* 3:395–400
- Lai D (2005) Monitoring the SARS epidemic in China: a time series analysis. *J Data Sci* 3:279–293
- Miller B, Kassenborg H, Dunsmuir W, Griffith J, Hadidi M, Nordin JD, Danila R (2004) Syndromic surveillance for influenzalike illness in an ambulatory care network. *Emerg Infect Dis* 10:1806–1811

- Montgomery DC (2005) Introduction to statistical quality control, 5th edn. Wiley, New York
- Morton AP, Whitby M, Mclaws M, Dobson A, Mcelwain S, Looke D, Stackelroth J, Sartor A (2001) The application of statistical process control charts to the detection and monitoring of hospital-acquired infections. *J Qual Clin Pract* 21:112–117
- Reis B, Mandl KD (2003) Time series modeling for syndromic surveillance. *BMC Med Inform Decis Mak* 3:2
- Rogerson PA, Yamada I (2004) Approaches to syndromic surveillance when data consist of small regional counts. *Morb Morality Wkly Rep* 53:79–85
- Tsui K-L, Chiu W, Gierlich P, Goldsman D, Liu X, Maschek T (2008) A review of healthcare, public health, and syndromic surveillance. *Qual Eng* 20:435–450
- Woodall WH (2006) The use of control charts in health-care and public-health surveillance. *J Qual Technol* 38:89–104
- Woodall WH, Marshall JB, Joner MD Jr, Fraker SE, Abdel-Salam A-SG (2008) On the use and evaluation of prospective scan methods for health-related surveillance. *J R Statist Soc A* 171(1):223–237
- Wu TSJ, Shih FYF, Yen MY, Wu JSJ, Lu SW, Chang KCM, Hsiung C, Chou JH, Chu YT, Chang H, Chiu CH, Tsui FCR, Wagner MM, Su IJ, King CC (2008) Establishing a nationwide emergency department-based syndromic surveillance system for better public health responses in Taiwan. *BMC Public Health* 8:18