N S A

NIGERIAN STATISTICAL ASSOCIATION

2007 CONFERENCE PROCEEDINGS

THEME:

STATISTICS: AN IMPERATIVE FOR GOOD GOVERNANCE.

Rasheed

VENUE:

NELROSE HOTEL, ASABA, DELTA STATE.

Pags 123 – 128

DATE:

4TH - 7TH SEPTEMBER 2007

TABLE OF CONTENTS

STATISTICAL MONITORING MODELS AND ITS APPLICATION TO MEDICAL ADMINISTRATION

R.A. ADEYEMI
Department of crop Production,
School of Agriculture and Agricultural Technology,
Federal University of Technology,
P.M.B. 65, MINNA- NIGERIA ,
Mobile +234-08035861447, email: adeyemira@yahoo.ca

ABSTRACT
Statistical models provide a unique description to available data from public health surveillance systems, which can provide meaningful measures of population risks for disease, disability, and death. Analysis and evaluation of these data help public health practitioners react to important health events in a timely manner both locally and nationally. This paper reviews some statistical models commonly used for detection of outbreaks of diseases epidemics and implemented on real life data. Two of the methods were considered for illustration. It was found that EWMA chart is slightly more sensitive than the corresponding CUSUM chart to detect a small shift from the process mean level.

KEYWORDS: *Dynamics Models, CUSUM. EWMA, Average run length, Acceptable Quality level, rejectable Quality level, Communicable diseases, Hospital data*

1. INTRODUCTION

Data from public health surveillance systems can provide meaningful measures of population risks for disease, disability, and death. Analysis and evaluation of these surveillance data help public health practitioners react to important health events in a timely manner both locally and nationally. Statistical methods such as outbreak detection methods are frequently used by epidemiologists and assist them in this task. The methods allow for the rapid assessment of changes in frequencies and rates of different health outcomes and for the characterization of unusual trends or clusters. The detection methods are used to enhance the public health practitioner's ability to identify and characterize unusual trends or clusters in public health surveillance data (for detail, see Langmuir (1963); Thacker *et.al* (1989); Teutsch *et.al* (2000)).

In this paper, it is important to note that data used for illustration of the methods are historical real life data caused by an infectious disease outbreak, and more important, they may or may not necessarily be from health sector, and data can come from simulation of a process or laboratory experiment.

Public health surveillance data do facilitate the timely detection of previously difficult to recognize outbreaks and allow the operators rapidly and promptly respond to outbreaks and take urgent action to savage any impending problem. Pattern recognition methods, which are not discussed in detail in this paper, are useful for identifying symptoms (or sets of symptoms) that deviate from the expected baseline and situation requiring attention. Visually display of information using charts and graphs are likely methods to identify random variations that are not epidemiologically important. A retrospective analysis can be desirable to provide diagnostic information about the process, which it is being monitored.

This paper reviews some statistical models commonly used for early detection of outbreaks of diseases epidemics and implemented on real life data. Two of the methods were used as an illustration for detection of such outbreaks.

3. STATISTICAL MODELS

3.1. Historical Limits Models

The methods can be used to compare the number of reported cases in the current (say 4-week) period for a given health outcome with historical incidence data on the same outcome from the preceding years (say 5 years) and is based on a comparison of the ratio of current reports with the historical mean.

The method is based on the formular

$x_o/\mu > 1 + (2^*\sigma_x/\mu)$ -- (1)

where x_o is the current total for a 4-week interval, μ is the mean of 15 totals of 4-week intervals (including the same 4-week period, the preceding 4-week period, and the subsequent 4-week period over the preceding 5 years of historical data), and σ_x is the standard deviation of these 15 historical incidence data values.

3.2 Log-Linear Regression Model

A log-linear regression model was developed Farrington *et al.* (1996). The method has often been used to analyze national-level infectious disease data reported to the Communicable Disease Surveillance Center (CDSC) in the United Kingdom.

The log-linear regression model are usually used to assess different types of variation in communicable disease incidence data is represented as

$\log \mu_i = \alpha + \beta t_i$ .... (2)

where μ_i is the mean of the baseline, α is the threshold value, βt_i is the systematic component of the model, and *i* is the week indicator.

The exceedance score *X* is estimated as, $X = (y_0 \,.\, \mu_0)/(Y \,.\, \mu 0) > 1$ .... (3) for counts considered epidemiologically significant (e.g., counts of 5 or more in the preceding 4 weeks), where y_0 is the current weekly count, μ_0 is the corresponding historical mean, and *U* is the expected acceptable shift based on a 2/3 power transformation.

3.3 Quality Control Cumulative Sums Methods

Cumulative Sum (CUSUM) chart was originally developed by Page (1954) for process control of variables, which are widely used in industry.

A standard CUSUM charts are often used to detect an upward or a downwards shift in process quality (one –sided CUSUM chart) or shift is in both directions (two-sided CUSUM chart). To monitor a positive shift from the goal value *k*, the CUSUM Statistic

$S_t = \max(0, x_t - k + S_{t-1})$ --- (4)

is used, while the statistic

$S_t = \min(0, x_t - k + S_{t-1})$ ------------------------------------ (5)

is used to detect a negative shift. The process is taken to be out of control if $S_t \geq h$ for an upward shift or $S_t \leq -h$ for a downward shift. In this work the starting values $S_0 = 0$ is taken to be zero.

The procedures for determination of the CUSUM k and h are discussed extensively in the literature (see, BSI 1982, Part IV page 1954; Lucas, 1985; Osanaiye et. al. 1989).

3.4 Quality Control Compound Smoothing Model

A quality control method called compound smoothing (4253H) was developed by Stern and Lightfoot.(1999).

The method use the formular as to compare the current count x_o with a threshold

$x_o > \beta + 2^*\sigma_x$ (6)

where β is the smoothed baseline, and σ_x is the standard deviation calculated as the differences between the smoothed value and the raw value for each data point.

3.5 Exponential Weighted Moving Average (EWMA) Model

The EWMA charting procedures are sometimes used to monitor the rate of occurrences of rare events where the time between two consecutive occurrences is exponentially distributed.

Suppose an EWMA chart is intended for detecting an increase in the exponential mean β and it is obtained by plotting

$Z_t = \max\{A, (1-\lambda_R)Q_{t-1} + \lambda_R X_t\}$ ---------------------------------------(7)

against t, for t = 1, 2, ... where λ_R is a smoothing constant such that $0< \lambda_R < 1$, A is a nonnegative boundary and $Z_0 = u$, $A\leq u \leq h_R$ An out of control signal is given at first t for which $Z_t \geq h_R$.

Similarly, a lower- sided EWMA chart is intended for detecting a decrease in the exponential mean β and is obtained by plotting

$$Z_t = \min \{B, (1-\lambda_L)q_{t-1} + \lambda_L X_t\} \quad \text{--(8)}$$

against t, for t = 1, 2, ... where λ_L is a smoothing constant such that $0< \lambda_L < 1$, B is a nonnegative boundary and $Z_0 = u$, $h_L \leq u \leq B$ A signal is given at first t for which $q_t \leq h_L$, (for detail procedure for design see Gan. 1998.)

3.6 Cyclical Regression Models

To assess periodic outbreaks of epidemics, Cyclical Regression models may be deemed desirable.

The cyclical linear regression model can be written as

$$Y = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 \mathrm{Cos}(2\pi t) + \beta_4 {*}\sin(2\pi t) \quad \text{---------------------------------- (9)}$$

where y represents total number of death reported from two simultaneous outbreaks (say pneumonia and influenza) for a particular week, β_0 is the intercept, β_1 and β_2 represent terms associated with the secular trends, and β_3 and β_4 represent cyclical terms associated with seasonal trends, and t, time sampling interval (e.g. weekly, monthly , quarterly)

3.7 Short-Term Methods.

The short-term implementation methods, or drop-in surveillance, are represented primarily by traditional quality control methods such as the P Chart, the moving average, and CUSUM. It may be desirable to calculate a 2 x 2 table chi-square since epidemiologists are more familiar with this summary statistic. For these methods, the average run length is usually 30 days. The length of time for the surveillance periods tends to be very short (approximately 21 days); therefore, seasonality factors are less important in the assessment of daily aberrations.

4 DATA APPLICATION

The data used for illustration of the methods in this paper was obtained as transcription from Research and Record Unit of Ahmadu Bello Teaching Hospital (ABUTH), Kaduna, Nigeria. It consists of 285 weekly sequences of reported cases of five infectious diseases (Malaria, Measles, Meningitis, pneumonia and Tuberculosis) between 1st January 1991 and 11th June 1996. These are extracted from a form containing 34 infectious diseases (including Food Poison and HIV/AID) tabulated by age and sex of which these five diseases are prominent (Adeyemi, 2006)

4.1 Detection of Outbreak of an epidemic (CUSUM chart)

Malaria (Case):

The current mean level = 3.34 approximately $\mu_a = 4$ with standard deviation $\sigma = 3.09$. Suppose the administration or authority demands a shift of $\boldsymbol{1.5\sigma}$ from μ_a to be the rejected level, then a shift of 1.5σ in the positive direction yields $\mu_d = 9$ (approximately).

Using equation (4), the CUSUM model to monitor outbreaks of malaria is :

$S_i = \max(0, X_i - 7 + S_{i-1}) \geq 7$; meaning that an out of control signal will be indicated whenever $S_i \geq h$

Table 1 shows the malaria data and corresponding CUSUM Statistics for the 3rd and 4th quarters of 1994 and implemented model to detecting the outbreak of an epidemic It shows that an out of control signal is first indicated at No. 14, which corresponds to 10th of January 1995. It indicates that it is statistically out of control with respect to the scheme, meaning an epidemic is detected.

Table 1: CUSUM tabulation for malaria data

Sample No.	X	X-k	$\sum(x-k)$	Si =max {0, $X_i - k + S_{i-1}$}
1	1	-6	0	0
2	0	-7	-7	0
3	2	-5	-12	0
4	1	-6	-18	0
5	1	-6	-24	0
6	1	-6	-30	0
7	2	-5	-35	0
8	18	11	-24	0
9	17	10	-14	0
10	5	-2	-16	0
11	4	-3	-19	0
12	4	-3	-22	0
13	15	8	-14	0
14	47	40	26	26*
15	43	36	52	52*
16	6	-1	51	51*

4.2 Detection of Outbreak of an epidemic (EWMA Chart) Malaria

A one sided-upper EWMA chart desired with an in control ARL of 800 (i.e. that is the scheme will give a false alarm with a probability of 0.00125), and optimal in detecting increase in counts rate of $\beta_1/\beta_0 = 6$, this yields the values of the parameters below:

$Q_0 = \beta_0 = 3$ $A = 0.5 \times \beta_0 = 1.5$, $\lambda_R = 0.295$, $h_R = 2.80 \times 3 = 8.40$

where λ_R is design parameter, $Q_0 = \beta_0$ and β_1 is the when process is in –control mean and out –of –control mean respectively, h_R is the limit (Threshold) value, A is the lower bound

Therefore, the EWMA model for monitoring outbreak of malaria s is;

$\mathbf{Z_t = Max\{ A, X\lambda_R + (1 - \lambda_R)Z_{t-1}, > h_R}$

This implies that an out of control signal will be indicated as soon as $Z_t \geq h_R$

Table 2 below shows the data for malaria for the 3rd and 4th quarters of 1995 and corresponding EWMA statistics to detect the outbreak of an epidemic

Table 2: EWMA tabulation for malaria data

Sample No.	X_i	$Z_t = \text{Max}\{ A, X\lambda_R + (1 - \lambda_R)Z_{t-1}$	$Z \geq 8.4$
1	1	3	
2	0	2.115	
3	2	2.081075	
4	1	1.762158	
5	1	1.537321	
6	1	1.378812	
7	2	1.562062	
8	18	6.411254	
9	17	9.534934	Out of control
10	5	8.197128	"
11	4	6.958976	"
12	4	6.086078	"
13	15	8.715685	"
14	47	20.00956	"
15	43	26.79174	"

An out of control signal is first indicated on sample No. 9, which corresponds to 6th December 1994.

5. SUMMARY

CUSUM and EWMA models were designed and implemented on some sample data, It was observed that both countable data CUSUM and Upper- sided EWMA charts designed had performed creditable and were sensitive in detecting the outbreaks of epidemics. In both all cases , the EWMA chart indicated an out of control signal on sample No. 9, which corresponds to 6th December 1994 while CUSUM chart indicated an out of control signal is at No. 14 , which corresponds to 10th of January 1995.

Thus, monitoring models are powerful tools, which can be used to obtain first hand information about an aberration in events of disease attack, disability or death. The EWMA and CUSUM schemes illustrated in this paper further confirms the vital role Quality Control techniques are capable of playing in a non-manufacturing sector as well.

6. CONCLUSION

The main goal in the development and implementation of these methods is to provide local and state health departments with a tool to assist in the best application of often-limited resources during epidemiological investigations of important public health events.

From the above analysis, it can be concluded that an EWMA chart is slightly more sensitive than corresponding CUSUM chart in detecting small or moderate increases in the process mean level; which is consistent with the findings of Moustakides (1986) and Gan (1998).

REFERENCES

Adeyemi R.A (2006)."Retrospective analysis and Design of CUSUM and EWMA charts for detection of outbreak of disease epidemics" unpublished M.Sc. graduate research project, Department of Statistics, University of Ilorin, Nigeria.

Thacker SB, Berkelman RL. (1988), Public health surveillance in the United States. *Epidemiol Rev.* ;10:164–190.

Teutsch SM, Churchill RE, eds(2000) *Principles and Practice of Public Health Surveillance.* New York, NY: Oxford University Press.

Centers for Disease Control and Prevention(1990). Guidelines for investigating clusters of health events. Appendix. Summary of methods for statistically assessing cluster of health events. *MMWR Morb Mortal Wkly Rep.*;39(RR-11):17–23.

Lucas JM (1985). Counted data CUSUM's. Technometrics 1985;27:129-44.

Gan, F.F. (1991) "An optimal design of CUSUM quality control charts," *Journal of Quality Technology,* **23**, 279- 286.

Osanaiye P.A. and Talabi O. (1989) On some non-manufacturing applications of counted data cumulative sum (CUSUM) control schemes, *The Statistician.*

Page, E.S. (1954) "Continuous Inspector Schemes, "*Biometrika*, **41**, 100-114.

Hutwagner LC, Maloney EK, Bean NH, Slutsker L, Martin SM. Using laboratory-based surveillance data for prevention: an algorithm for detecting *Salmonella* outbreaks. *Emerg Infect Dis.* 1997;3:395–400.

Stroup DF, Williamson GD, Herndon JL, Karon J. Detection of aberrations in the occurrence of notifiable diseases surveillance data. *Stat Med.* 1989;8:323–329.

Farrington CP, Andrews NJ, Beale AD, Catchpole MA. A Statistical algorithm for the early detection of outbreaks of infectious disease. *J Roy Stat Soc A.* 1996;159:547–563.

Simonsen L, Clarke JM, Stroup DF, Williamson GD, Arden NH, Cox NJ. A method for timely assessment of influenza- associated mortality in the United States. *Epidemiology.* 1997; 8:390–395.

Stern L, Lightfoot D. Automated outbreak detection: a quantitative retrospective Analysis.*Epidemiol Infect.* 1999;122:103–110.

Hutwagner L, Thompson W, Groseclose S, Williamson GD. An evaluation of alternative methods for detecting aberrations in public health surveillance data. *American Statistical Association 2000 Proceedings of the Biometrics Section*. 82–85.
Stoumbos ZG, et.al. (2000). *The state of statistical process control as we proceed into the 21st century. J Am Stat Assoc.*;95:992–998.