

Statistical Quality Control; a Tool for Monitoring Epidemic Diseases Outbreak

¹O.J. Braimah, ²P.E. Omaku, ³Y.K Saheed and ⁴A.A. Momo

¹Department of Statistics, Abdur-Raheem College of Advanced Studies, Igbaja (An Affiliate of Al-Hikmah University, P.M.B 1601) Ilorin, Kwara State, Nigeria

²Department of Mathematics and Statistics, Federal Polytechnic Nasarawa, Nasarawa State, Nigeria

³Department of Computer Science, Abdur-Raheem College of Advance Studies Igbaja (An Affiliate of Al-Hikmah University), P.M.B 1601, Ilorin, Kwara State, Nigeria

⁴Department of Statistics, University of Ilorin, P.M.B 1515 Ilorin, Kwara State, Nigeria

Abstract: This study is aimed at designing CUSUM and EWMA control chart and comparing their performances. The data used in this paper are records of outbreak of some diseases collected from Edo State Civil service Hospital, Benin City and are therefore secondary data. The data covered the period, January 2006 to December 2010. Cumulative Sum (CUSUM) and Exponentially Weighted Moving Average (EWMA) control charts scheme were designed and used to study the rate of outbreak of some diseases. Finally, the performances of the two designed control chart schemes were compared. Both CUSUM and EWMA were capable of detecting shifts in diseases data. EWMA charts tended to be slightly slower in providing alarms, but were much easier to set up. Overall CUSUM charts are more efficient to use. Use of these techniques could allow detection of changes in time to mitigate the recurrent pattern of the outbreak of epidemics.

Keywords: Cumulative Sum (CUSUM), Decision Interval (DI), Exponentially Weighted Moving Average (EWMA), Lower Control Limit (LCL), Upper Control Limit(UCL), Acceptable Quality Level (AQL), Rejectable Quality Level (RQL), In-control Average Run Length (ICARL), Out-of-Control Average Run Length (OO-ARL), Average Run Length (ARL), Run Length (RL)

I. Introduction

The use of techniques and activities to achieve, sustain and improve the quality of a product or service is quality control. Today, there is no single universal definition of quality. Some people view quality as “performance to standards” others view it as “meeting customer’s needs” or “satisfying the customer”.

Furthermore, quality is reviewed in using different approaches by the quality gurus like Philip B. Crosby, Joseph .M. Juran, G. Tagushi E, Deeming, and other sources gave quality definition but with a convergence (goal), as discussed below: W. Edwards Deming stated that quality should be aimed at the views of the customers present and future needs. Philip B. Grosby defined quality as conformance to requirement. Joseph M. Juran described quality as “fitness for purpose”, that is a product or service should do exactly what it is designed to do. Genichi Taguchi gave his own definition of quality as the degree of variation from target values. ISO- viewed quality as the totality of features and characteristics of a product or service that bears on ability to meet stated and implied needs.

So, quality assurance of health care delivery requires proper monitoring of the patients that patronize their service, apart from other means of studying the trend of the ugly situations they find themselves in – outbreak of prominent diseases epidemic, whether informally via a system such as peer review. Current methods of studying the trend are, however, subjective or implicit. More objective methods of monitoring the patients affected by the outbreak of these prominent diseases and studying the point of change in the recurrent pattern of such diseases based on statistical process control techniques like cumulative sum (CUSUM) and Exponentially Weighted Moving Average (EWMA) as discussed in this paper. To help the prospective health managements make predictions for future outbreak using historical data on outbreak of common diseases, this works well if the future outbreak is properly managed or corrected using the historical data. This is done by the use of by the use of Cumulative Sum (CUSUM) and Exponentially Weighted Moving Average (EWMA) control charts. Also, certain changes in the data must be properly noted because the prediction made can be error leading, it may be giving increasing outbreak at a particular period of time where as it is not true. So, it is very important for health management to know if there are changes in the number of reported cases. While management wants to know quickly about a change, they do not want false alarms. The tragedy between fast response to changes and the number of false alarms is a key design consideration for any monitoring system and is discussed in detail in this paper. In order to monitor any process, it is very important to know what data is available and representative of the process. So historical data collected form the Data Record Department of Edo State Civil Service Hospital is suitable for this study, thus, this data was chosen as the data element to be analyzed.

II. Materials And Methods

2.1 The Cusum Technique

It may be difficult to see small changes in the distribution of x_i on a Shewhart chart. It can be easier to see those changes if one “accumulates” deviations from some standard value for x_i . A “raw” Cumulative Sum sequence is defined by: $C_i = (x_i - k) + C_{i-1}$.

In statistical quality control, the **CUSUM** (or **cumulative sum control chart**) is a sequential analysis technique proposed by E. S. Page (1954) of the University of Cambridge. It is typically used for monitoring change detection.

As its name implies, CUSUM involves the calculation of a cumulative sum (which is what makes it “sequential”). Samples from a process x_i are assigned weights k , and summed as follows:

$C_0 = 0$, if the value is negative. So we plot either $C_i = \max(0, x_i - k + C_{i-1})$ or $C_{i+1} = \max(0, C_i + x_i - k)$ against i .

When the value of C exceeds a certain threshold value ($>h$), a change in value has been found. The above formula only detects changes in the positive direction. When negative changes need to be found as well, the min operation should be used instead of the max operation, and this time a change has been found when the value of C is below the (negative) value of the threshold value ($<-h$). As a means of assessing CUSUM's performance, Page defined the average run length (A.R.L.) metric; “the expected number of items sampled before action is taken

Summary Of Cummulative Sum (Cusum) Chart:

CUSUM chart Originally proposed by: E. S. Page(1954)

Process Observations: Rational subgroup size $n = 1$

Measurement type: Cumulative sum of a quality characteristic

Quality characteristic type: Count data

Underlying distribution : Poisson Distribution

Size of shift to detect: 1.0σ

Process mean chart Center line: The target value T of the quality characteristic.

Control limits are:

Upper control limit(UCL): $C_i^+ = \text{MAX} [0, x_i - K + C_{i-1}^+] > h$, for positive change.

Lower control limit(LCL): $C_i^- = \text{MIN} [0, x_i - K + C_{i-1}^-] < -h$, for negative change.

Plotted statistic:
$$C_i = \sum_{j=1}^i \bar{x}_j - T$$

Where, T - the target mean of the process, K - Reference value for simultaneous low and high side scheme, h - the decision interval of CUSUM, i - the i^{th} subgroup, j - the subgroup items. All of K , C_i^+ and / or C_i^- are (appropriately chosen) parameters of the scheme.

2.2 Exponentially-Weighted Moving Average(Ewma) Chart

“If the CUSUM chart is the sum of the entire process history,

May be a weighted sum of the recent history would be more meaningful”

The use of this control chart was first published in the article “Control chart tests based on geometric moving averages” (Roberts, 1959). While Shewhart charts only consider the most recent data point in testing to determine if statistical limits have been exceeded, EWMA charts consider all previous points using a weighing factor that makes the outcome more influenced by recent points.

In statistical quality control, the EWMA chart (or exponentially-weighted moving average chart) is a type of control chart used to monitor either variables or attributes-type data. While other control charts treat rational subgroups of samples individually, the EWMA chart tracks the exponentially-weighted moving average of all prior samples. EWMA weights samples in geometrically decreasing order so that the most recent samples are weighted most highly while the most distant samples contribute very little. Although the normal distribution is the basis of the EWMA chart, the chart is also relatively robust in the face of non-normally distributed quality characteristics. There is, however, an adaptation of the chart accounts for quality characteristics that are better modeled by the Poisson distribution. The chart monitors only the process mean. Also EWMA Charts are generally used for detecting small shifts in the process mean. They will detect shifts of 0.5 sigma to 2 sigma much faster than Shewhart charts with the same sample size. In brief, after multiplication by a weighting factor w , the current measurement is added to the sum of all former measurements, which is weighted with $(1 - w)$. Thus, at each time t ($t = 1, 2, \dots$), the test statistic z_t , $z_t = [w \bar{x}_t + (1 - w)z_{t-1}]$, with $w [0;1]$, can be obtained. The computed z_t values are displayed on a control chart over the course of time. The EWMA control chart requires selection of two parameters before setup:

1. The first parameter is w , the weight given to the most recent rational subgroup value. So, w must satisfy $0 < w \leq 1$, but selecting the "right" value is a matter of personal preference.

- The second parameter is sigma limit q, the multiple of the rational subgroup standard deviation that establishes the control limits.

But, the selection of w and q according to Neubauer(1997) in the article “The EWMA control chart: properties and comparison with other quality-control procedures by computer simulation,” described a graphical method for determining the ARL for an EWMA control chart was used in this research work .

Summary Of Ewma Control Chart

EWMA chart Originally proposed by: S. W. Roberts(1959)

Process observations: Rational subgroup size n = 1

Measurement type: Moving average of the quality characteristic

Quality characteristic type: Count Data

Underlying distribution : Poisson distribution

Size of shift to detect: 1.0σ

The target value of the quality characteristic is denoted as: Z₀

Control limits: The target or center line for the control chart is the average of historical data. The upper (UCL) and lower (LCL) limits are:

$$UCL = \mu_0 + q\sigma_0 \sqrt{\frac{w}{(2-w)n}} \quad \text{and} \quad LCL = \mu_0 - q\sigma_0 \sqrt{\frac{w}{(2-w)n}}$$

Plotted statistic: z_t = [w x̄_t + (1 - w)z_{t-1}]

Where, q – the sigma limit value, σ₀ – historical data standard deviation., w – the weighting or smoothing value, n – the subgroup size, μ₀ - the historical mean or mean of the historical data, z_t – the EWMA statistic, z_{t-1} – the previous EWMA statistic

x_t – the observed values **COMPUTATION OF CONTROL LIMIT**

By default, the central line on an EWMA chart indicates an estimate for μ, which is computed as:

$$\hat{\mu} = \bar{X} = \frac{n_1\bar{X}_1 + \dots + n_N\bar{X}_N}{n_1 + \dots + n_N}$$

The default limits are computed as q = 2.30 (referred to as 2.30σ limits).

The upper and lower control limits are

$$UCL = \bar{X} + q\sigma \sqrt{\frac{w}{n(2-w)}}$$

$$LCL = \bar{X} - q\sigma \sqrt{\frac{w}{n(2-w)}}$$

where ‘n’ is the subgroup size.

Two characteristics of the control charts were evaluated:

- How effective were each of the control chart at detecting real shifts with limited numbers of false alarms?
- How efficient were each of the control charts as gauged by the amount of effort it required to set up and produce each chart

2.3 Analysis Techniques Used

In order to compare the effectiveness of the CUSUM and EWMA methodologies to detect changes in underlying distributions, sample data sets were chosen for analysis, namely, data sets on malaria, hypertension and diarrhea.

Method of Analysis

CUSUM and EWMA charts were generated using Minitab for the data sets on the selected diseases.

The Following Charts were Plotted:

The first chart displays the cumulative malaria’s monthly recording over the period of analysis and was generated using **Excel 2007**. This provided a visual depiction of the disease pattern over time.

The second chart displays a CUSUM chart as generated by **Minitab 13** for malaria using h and k values as determined using **Anygeth.exe**. For the CUSUM charts, the in-control ARL was set to 15 months. As such, one would expect to see about one false alarm every 15 month. This says that when a process is in control one expects an out-of-control signal (false alarm) each 15 runs. That is, the scheme will give a false alarm with probability of 1/15.

The third chart displays a EWMA control chart for the data on malaria with ARL of 15 months. Similarly, one would expect to see one false alarm every 15 month. By examining the CUSUM and the EWMA chart, and comparing them to the shifts in the cumulative monthly recording chart generated by **Excel 2007**, we will assess the efficiency and effectiveness of each method relative to its ability to determining shifts in the outbreak without excessive false alarms being present.

2.3.1 Validation of CUSUM Methodology and Modification of Variables between Anygeth.exe and Minitab

When attempting the CUSUM validation phase of this effort against the malaria data set using h and k values obtained from Anygeth.exe within the Minitab program, several problems arose. This section will discuss the steps necessary to successfully integrate the two programs. Before one can determine the parameters of the CUSUM scheme, one must first select an in-control ARL. Recall that this specifies the average number of time periods until one gets a false alarm while in control. Longer ARLs imply fewer false alarms, but can result in a longer delay until a true alarm is signaled. The second parameter needed is to select the out-of-control mean. This is set based on the context of the problem, and is generally based upon the size of a change that starts to have serious practical effects. Larger departures are more robust against model misspecification (i.e., the distribution is not exactly as assumed). Using anygeth.exe, one can find the values of h and k that are tuned for the size of the departure desired. The book on Cumulative Sum Charts and Charting for Quality Improvement (Hawkins & Olwell, 1997) gave a complete discussion on the methodology of best choosing these variables as used in this paper. A program that operates under Windows called Anygeth.exe will generate h and k values for given values of ARL- in-control mean and out of-control mean. **While h and k are reasonably easy to calculate for a normal distribution, they are more complex for Poisson and other distributions. The use of the software greatly simplifies the task.**

Minitab and anygeth.exe use slightly different parameterizations, so one must convert between them. The following instructions should be followed carefully:

1. Calculate h and k using Anygeth.exe for the desired ARL and out of control state. From Figure 4 we obtained $h = 2.348$ and $k = 1.348$.

The values for h and k are selected to have a good balance between type 1 and type 2 error probabilities.

2. Determine the value of adjusted $k = \text{unadjusted } k - \text{in-control-mean } (k_0)$. But we cannot use the unadjusted value $k = 1.348$ from anygeth.exe .OR using $k = (k_1 + k_0)/2$. Therefore, for this work, we got adjusted $k = 1.348 - 1.153 = 0.195$.

2.3.2 Cusum Method

The Decision Interval(DI) form of the CUSUM to test for an upward shift in mean is set up as follows as discussed by Hawkins and Olwell (1997):

$$C_0^+ = 0,$$

$$C_i^+ = \text{MAX} [0, x_i - K + C_{i-1}^+].$$

This produces an alarm showing an upward shift in mean if $C_i^+ > h$.

$$C_0^- = 0,$$

$$C_i^- = \text{MIN} [0, x_i - K + C_{i-1}^-]$$

This produces an alarm showing a downward shift in mean if $C_i^- < -h$.

Determining the correct values of h and k to use is essential in creating a CUSUM system that will set off an alarm quickly when the mean has shifted without excessive false alarms during in control operations. The average time between alarms is termed the Average Run Length (ARL). Two ARLs are generally used, one for in-control and one for out-of-control at a specified level. The ARL is a function of h and k and the underlying distribution (in-control or out-of-control) of data.

A program that operates under Windows called Anygeth.exe will generate h and k values for the given values of ARL- in-control mean and out of-control mean.

Minitab 13, a commercially available statistics program will develop Decision Interval (DI) CUSUM charts using the values obtained from Anygeth.exe as used in this thesis.

2.3.3 Choice Of The Reference Value K

In the (upper) one-sided, univariate CUSUM procedure, the reference value K is often taken to be the average of the expected value of the process mean under

$H_0: K = k_0$ and $H_1: K = k_1$, where k_0 is the in-control mean known as the Acceptable Quality Level (AQL) – the on-target state and k_1 is the out-of-control mean known as the Rejectable Quality Level(RQL) – the off-target (unacceptable) state.

In order to accurately compare EWMA and CUSUM techniques, they must be set up on a “level playing field.” Both must be designed with the same ARL, or the results cannot be compared fairly. So, the methodology to set up the Minitab EWMA control chart function to have an ARL of 15 will be described and validated against the data set for malaria.

Neubauer (1997) in the article “The EWMA control chart: properties and comparison with other quality-control procedures by computer simulation,” described a graphical method for determining the ARL for an EWMA control chart. As seen in Figures 1 and 2 below, one can obtain the weighting factor “w” as a function of the desired shift sensitivity of the chart. For a shift of 1.0 standard deviation, a weighting factor of 0.15 can be obtained from the chart as shown in figure 1. And we set ARL at 15, using figure 2, the sigma limit (q) was obtained to be 2.30.

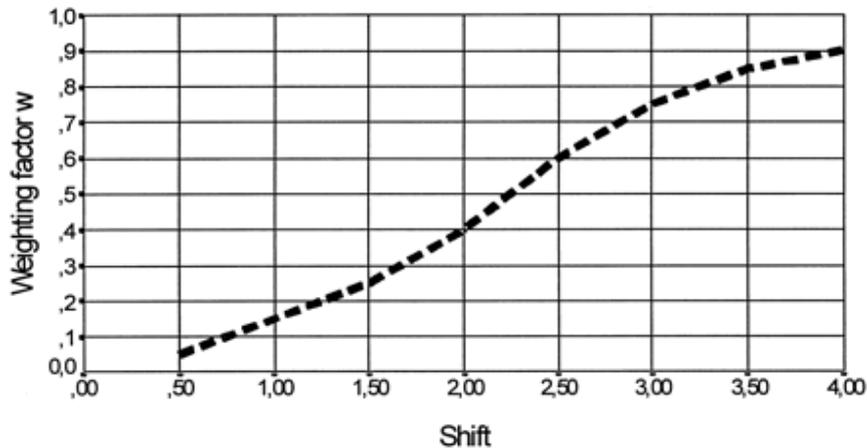


Figure 1: Optimal w for EWMA charts according to the shift d from Neubauer (1997)

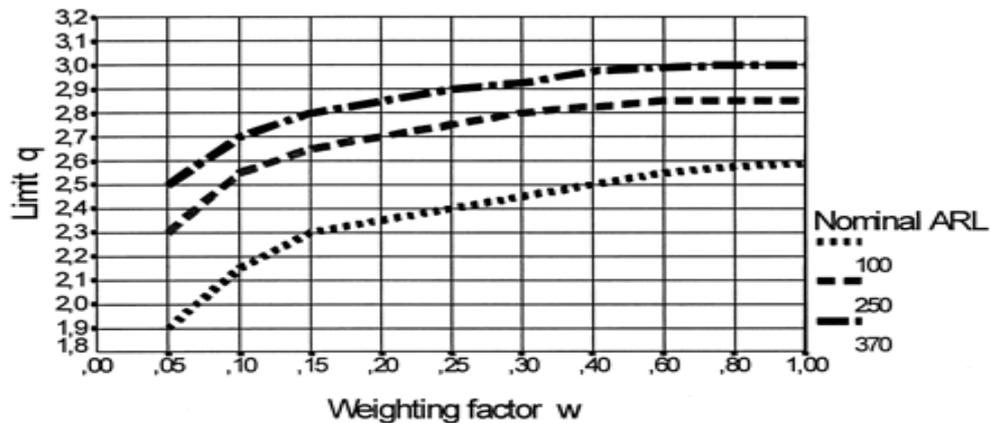


Figure 2: Determining the limit q of the EWMA chart after selection of w and the nominal ARL .

III. Data analysis:

The analysis of the data (record of outbreak of some diseases) obtained from Edo State Civil service Hospital, Benin City for period of January 2006 to December 2010 were considered using the two control chart schemes. Table 1 and 2 display the performances of the schemes, figure 3 depict how to select the Reference Value **k** and Decision Interval **h** and figure 5 to 14 depict the control charts for the three diseases respectively.

Table 1: Cusum And Ewma Charts Evaluation Using Malaria.

ARL	METHOD	FIRST ALARM DETECTED
15	CUSUM	8
	EWMA	24
10	CUSUM	6
	EWMA	24
5	CUSUM	4
	EWMA	24

Table2: Cusum And Ewma Chart Evaluation Using Some Prominent Diseases

DISEASE	METHOD	FIRST ALARM DETECTED WITH ARL 15
MALARIA	CUSUM	8
	EWMA	24
HYPERTENSION	CUSUM	5
	EWMA	40
DIARRHOEA	CUSUM	2
	EWMA	11

Figure 3: Cummulative Malaria Pattern For The Years 2006 - 2010

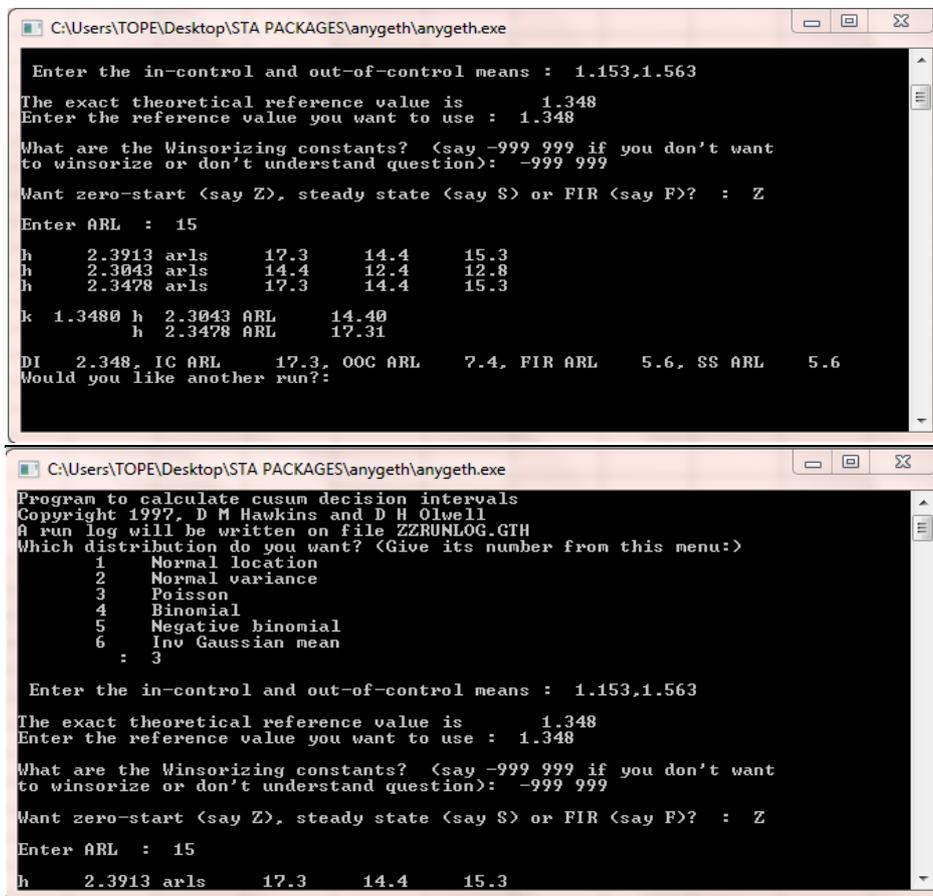
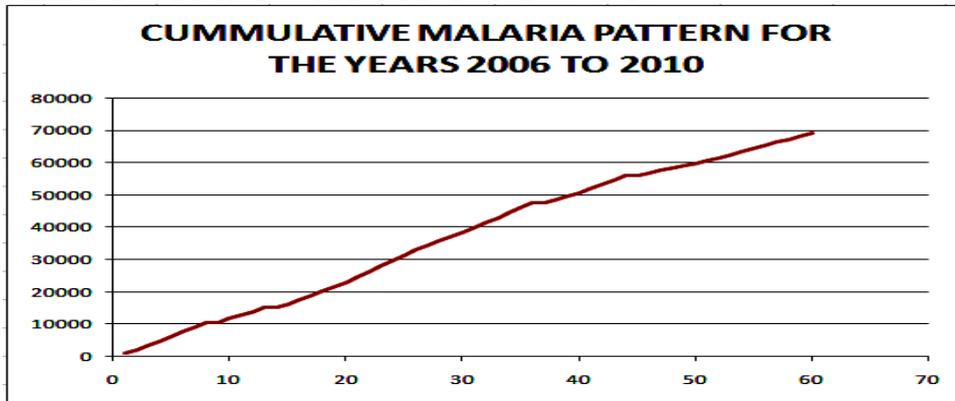


Figure 4: Anygeth.Exe Dialog For Determining Cusum Decision Intervals (Di)

The figure below displays the CUSUM chart for malaria using a K value of 0.195, and h value of 2.348 as determined using Anygeth.exe with an ARL of 15. The in-control mean (k_0) = 1.153 and the out-of-control mean (k_1) = 1.563. So, for k_0 and k_1 in thousand, we got 1153 and 1563 respectively.

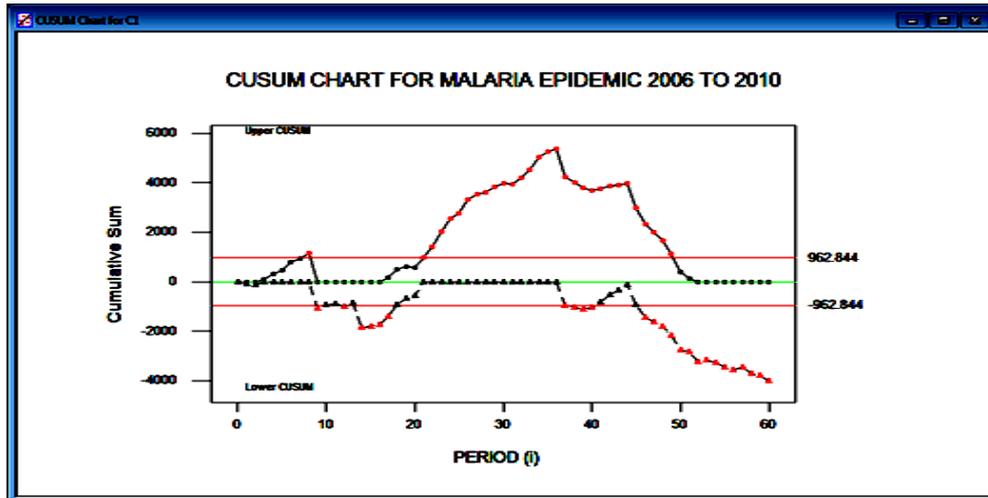


Figure 5: Cusum Chart For Malaria Epidemic 2006 To 2010 With Arl 15

The figure below displays the EWMA for malaria with an ARL of 15, weight of 0.15, and control limits at 2.30 standard deviations.

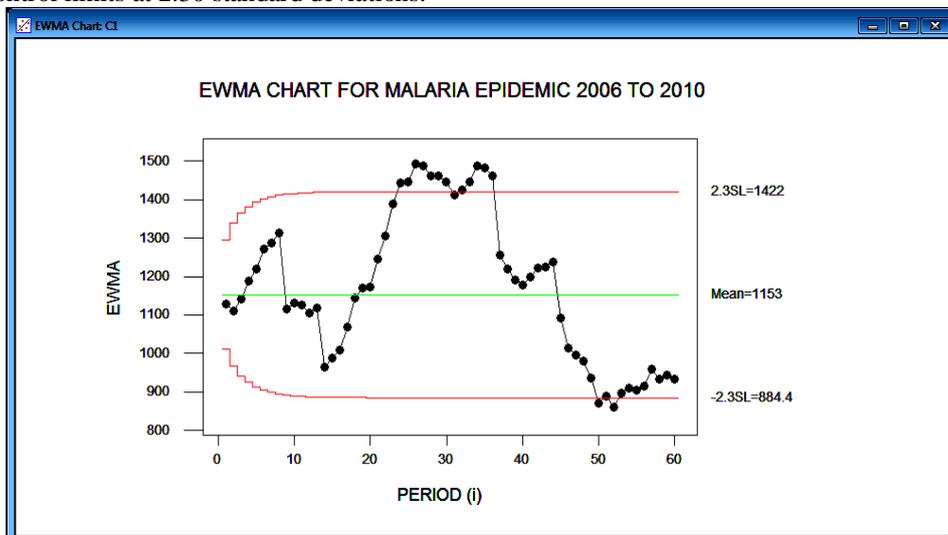


Figure 6: Ewma Chart For Malaria Epidemic 2006 To 2010

The figure below displays the CUSUM chart for malaria using $k_0 = 1.153$, $k_1 = 1.563$, h value of 1.696, K value of 0.195, IC ARL 12.2 and OOC ARL 5.9 as determined using Anygeth.exe.

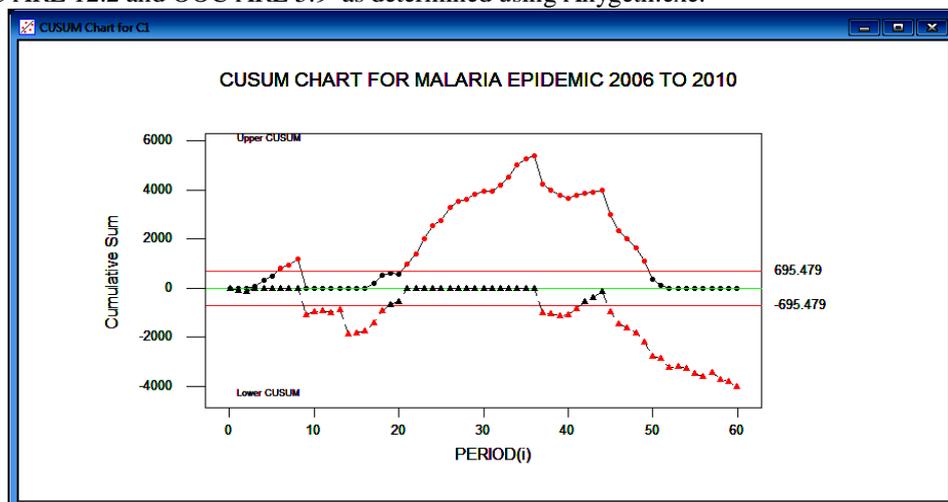


Figure 7: Cusum Chart For Malaria Epidemic With Arl Of 10

The figure below displays the CUSUM chart for malaria using $k_0 = 1.153$, $k_1 = 1.563$, h value of 0.696, K value of 0.195, IC ARL 6.4 and OOC ARL 3.8 as determined using Anygeth.exe.

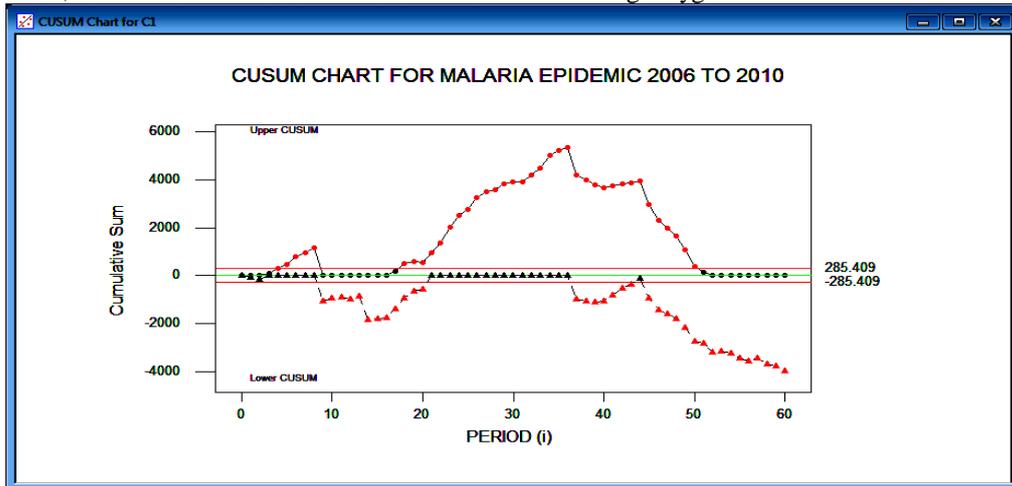


Figure 8: Cusum Chart For Malaria Epidemic With Arl Of 5

The below figure displays the cumulative number of hypertension patients plotted on the y axis versus months on the x axis.

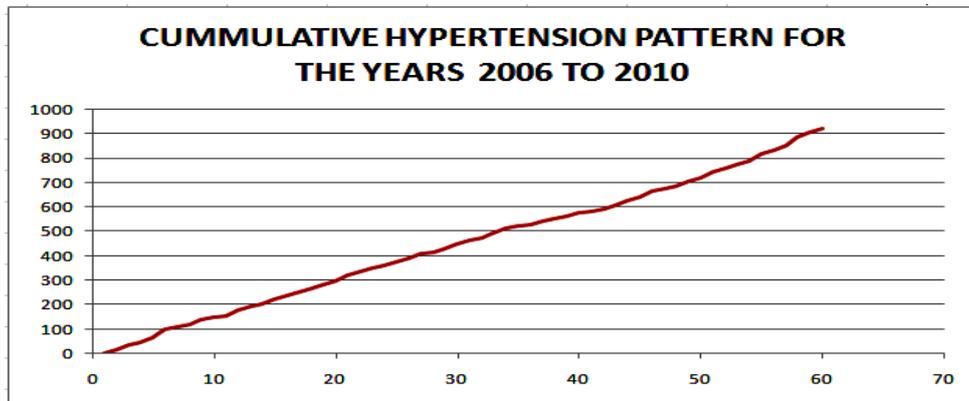


Figure 9: Cummulative Hypertension Pattern 2006 - 2010

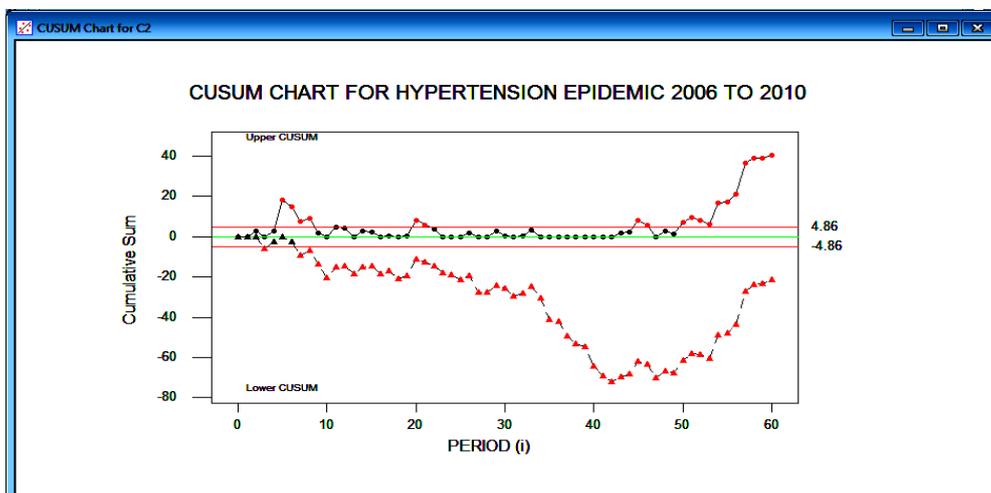


Figure 10: Cusum Chart For Hypertension Epidemic 2006 – 2010

Figure below displays the EWMA for hypertension with historical mean value of 16 and standard deviation of 5.374 and ARL of 15, weight (w) of 0.15, and control limits(q) at 2.30 standard deviations.

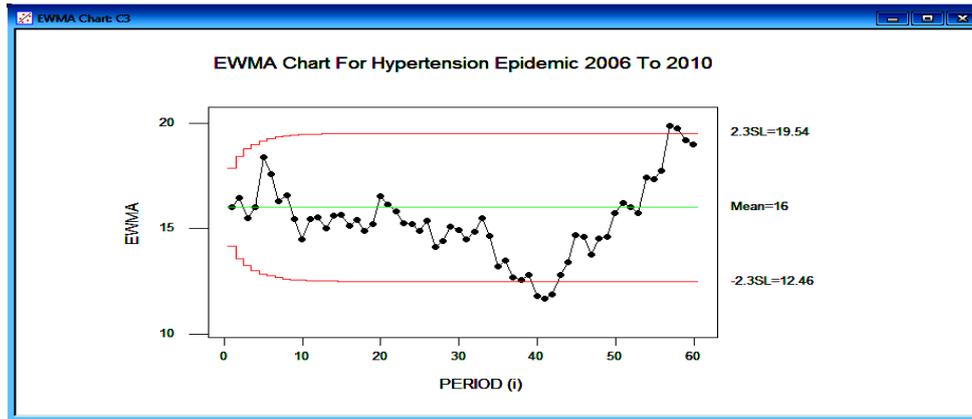


Figure11: Ewma Chart For Hypertension Epidemic 2006 To 2010

The figure below displays the cumulative number of patients affected by diarrhea plotted on the y axis versus months on the x axis.

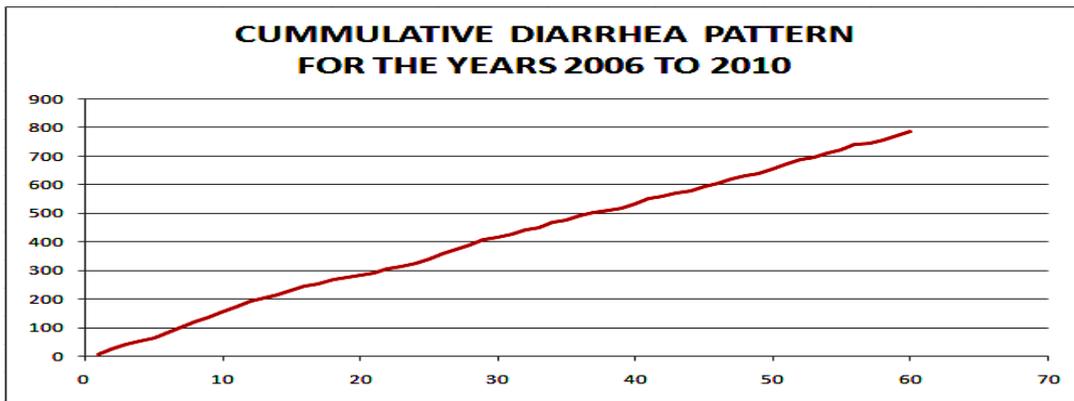


Figure 12: Cummulative Diarrhea Pattern For The Years 2006 - 2010

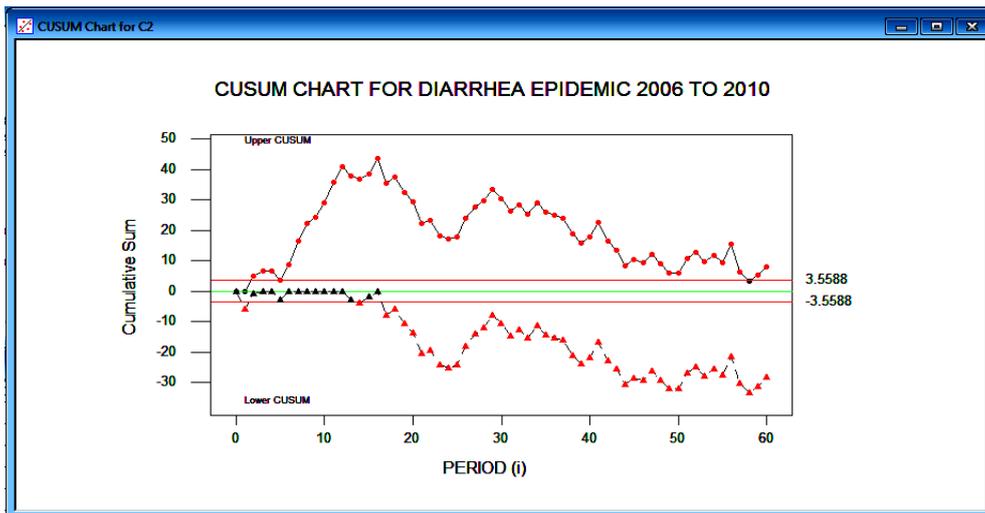


Figure 13: Cusum Chart For Diarrhea Epidemic 2006 – 2010

The figure below displays the EWMA for diarrhea with historical mean 13 and standard deviation of 4.129, ARL of 15, weight (w) of 0.15, and control limits(q) at 2.30 standard deviations.

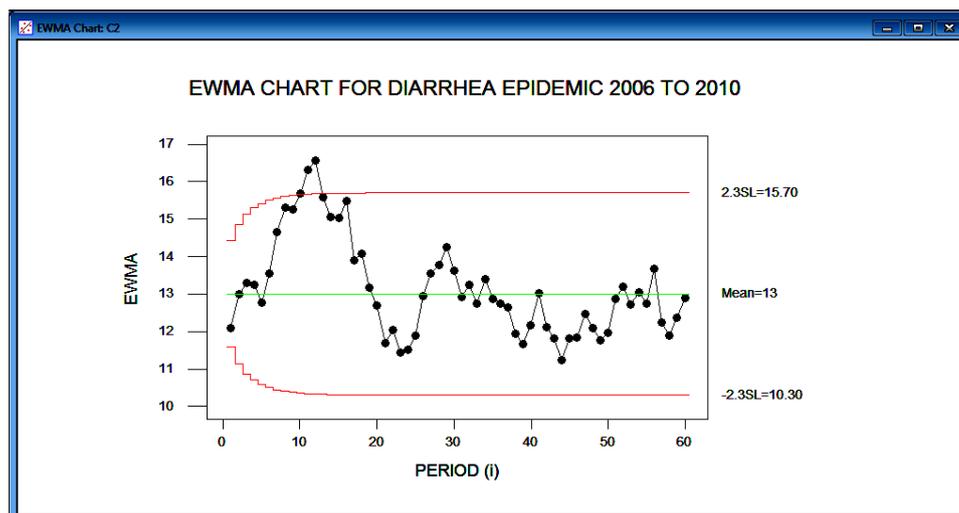


Figure 14: Ewma Chart For Diarrhea Epidemic 2006 To 2010

IV. Results

Generally it was observed that shorter ARL caused CUSUM to detect shift faster and longer ARLs caused longer delay of true alarm, but this did not have effect on EWMA irrespective of the ARL values set, we got alarm at the same point 24 for the three ARLs considered, that is 15, 10 and 5 using dataset on malaria as shown in table 1.

The effectiveness of the charts to determine shifts in underlying pattern and the efficiency of the charts relative to the time and expertise required to generate the charts were compared. Both methods were very capable at detecting changes in the disease recurrent (persistent) pattern when properly configured within Minitab. And we strongly observed that CUSUM generally provided a slightly faster alarm even according to the evaluation done in table 2 and figure 5 to 14. It was ascertained using datasets on some prominent diseases considered with ARL of 15 and we got the following results: for malaria, the CUSUM and EWMA produced the first alarm signal at points 8 and 24, hypertension: 5 and 40, diarrhea: 2 and 11. CUSUM detected alarms very quickly in three diseases considered, that is malaria, hypertension and diarrhea while EWMA still produced alarms fast slightly in two of the diseases, that is, diabetes and typhoid, but not as fast as that of CUSUM. Also EWMA was much easier to use and it requires less set up time than CUSUM, but CUSUM requires a custom set up for each of the prominent disease considered with an understanding of the underlying data distribution, EWMA is insensitive to the underlying data distribution.

V. Conclusion

Conclusively, since the CUSUM methodology detects shift more quickly for this type of dataset, it may be a more efficient choice for monitoring it. So, for most outbreaks of epidemic applications within the health organization CUSUM can be a more efficient tool in monitoring shift in the process mean.

References

- [1]. Barnard, G.A. (1959). "Control charts and stochastic processes". Journal of the Royal Statistical Society.
- [2]. Bower, K.M. (October 2000), Using Exponentially Weighted Moving Average (EWMA) Charts, Asia Pacific Process Engineer.
- [3]. Cembrowski G.S, Westgard J.O, Eggert A.A, Toren E.C, Jr. (1975): Trend detection in control data, optimization and interpretation of Trigg's technique for trend analysis.
- [4]. Crowder S.V (1987): A simple method for studying run-lengths of exponentially weighted moving average control charts. *Technometrics*; 29:401-407.
- [5]. Crowder S.V (1989): Design of exponentially weighted moving average schemes. *J Qual Technology*; 21:155-162.
- [6]. Deeming, W. Edwards(1986)- Out of crisis. Cambridge, Mass: MIT centre for Advanced Engineering Study.
- [7]. Evans James R., and William M. Lindsay(1999). *The Management and control of quality* 4thed.
- [8]. Grigg et al.; Farewell, V.T; Spiegelhalter, D.J (2003). "The Use of Risk-Adjusted CUSUM and RSPRT Charts for Monitoring in Medical Contexts". *Statistical Methods in Medical Research* 12 (2): 147-170.
- [9]. Grosby, Philip(1984), *Quality without Tears: The Art of Hassle - Free Management*, New York: McGraw - Hill.
- [10]. Garving, David A(1987): "competing on the Eight Dimensions of Quality", *Harvard Business Review*.
- [11]. Hawkins & Olwell(1997): "Cumulative Sum Charts and Charting for Quality Improvement".
- [12]. Hawkins & Olwell(1997), Anygeth.exe: <http://www.stat.umn.edu/cusum/software.htm>.
- [13]. Microsoft® Encarta® 2009. © 1993-2008 Microsoft Corporation.
- [14]. Montgomery, Douglas (2005). Introduction to Statistical Quality Control. Hoboken, New Jersey: John Wiley & Sons, Inc. ISBN 9780471656319.
- [15]. Neubauer(1997): Article on "The EWMA control chart: properties and comparison with other quality-control procedures by computer simulation".