*Journal of the Nigerian Association of Mathematical Physics*
*Volume* 14 (May, 2009), pp 171 – 176


# Detection of cholera outbreaks: A statistical scheme for its detection

**[1]A. O. Isere, [2]J. E Osemwenkhae and [3]I. W. Edokpa**
*[1,3]Department of Mathematics and Statistics, Ambrose Alli University, Ekpoma, Nigeria*
*[2]Department of Mathematics, University of Benin, Benin City, Nigeria*

***Abstract***

***In this paper, cumulative sum (CUSUM) control chart schemes are designed to detect the increase or decrease in the events of cholera outbreaks in Nigeria. The designed schemes are applied on the WHO weekly epidemiological record of cholera cases in Nigeria between 1996 and 2005 to demonstrate the application of the technique. From the demonstration, the technique has a good potential as a tool for detecting cholera outbreaks.***

***Keywords***
Cholera cases, WHO data, CUSUM, Detection, Outbreak.

## 1.0 Introduction

Cummulative sum (CUSUM) chart was conceived by Page [1] to give a tighter process control as against the classical scheme developed by Shewhart in 1920 – see Adeyemi [2]. The later was used to detect large shifts in process level. On the other hand, CUSUM does not only detect small shifts in process level but also does this easily and quickly. It has also been found to be more meaningful graphically - see and Lucas and Crossier [3].

Cumulative sum chart schemes are widely used in the industrial sector. They have recently found application in the non-manufacturing sectors such as in the area of detecting epidemics. While they are widely used for variable data, the applications to counted data have been quite interesting. The procedure for the design and implementation of the counted data CUSUM was developed by Lucas [4] who demonstrated that they are similar to that of the variable data. Osanaiye and Tolabi [5] applied CUSUM counted data on the diabetics disease data obtained from Ibadan.

The principal feature of the CUSUM control chart schemes is that successive values of a variable are compared with a predetermined target or reference value, $k$. The cumulative sum of the deviation from $k$ is given by:

$$S_N = \sum_{i=1}^{N} (Y_i - k) \qquad (1.1)$$

This is then plotted on a chart or tabulated. To monitor the shift from target, the CUSUM statistics

$$S_i = \max(0, Y_i - k + S_{i-1}) \qquad (1.2)$$

$$S_i = \max(0, k - Y_i + S_{i-1}) \qquad (1.3)$$

**[1]Corresponding author:**
**[1]abedris@yahoo.com**
**[2]Telephone: 08037378121, 08071066580**

are used for positive and negative shifts respectively. The process is taken to be out of control if $S_i \geq h$ for an upward shift or $S_i \leq -h$ for a downward shift (where $h$ is the decision interval).

This paper applies the design of a counted CUSUM control chart on World Health Organization (WHO) weekly epidemiology record of reported cases of cholera in Nigeria for ten years (1996 – 2005).

Generally, it is recognized that CUSUM control scheme is a sequence of Wald sequential probability Ratio Test (SPRT's) which allows for the development of its optimal property. Lucas and Crossier [3] went further to include the fast initial response (FIR) feature. The FIR features give simple procedure for detecting an out-of-control situation at start up more quickly. A signal is given whenever an outbreak is detected either with increasing or decreasing cholera cases.

## 2.0 Designs of outbreaks CUSUM scheme

The design procedure for the outbreak CUSUM involves the choice of the reference value, $k$ and the decision interval $h$. The parameter $h$ is primarily determined by the desired Average Run Length (ARL), that is, the acceptable frequency of false out-of-control signals. After the value of $k$ is determined, Table 1 (See Appendix) is used to choose the appropriate value of $h$.

The parameter $k$ is the reference value of the CUSUM. It is also referred to as the target value for the process for counted data CUSUM. It can be described as the goal value for the process, which is usually chosen between the acceptable process mean value, $\mu_a$ and the mean that CUSUM scheme should detect quickly, $\mu_d$, otherwise known as the barely tolerable mean value. Kemp [6] gave the expression for determining $k$ as

$$k = \mu_a + \tfrac{1}{2}\Delta \tag{2.1}$$

where $\Delta$ is the deviation from the acceptable process. Lucas et al [3] also recommended that the value of *k be* $\Delta/2$ for a scheme designed to detect a specific mean shift of $\Delta$, and it should also be chosen close to

$$(\mu_d - \mu_a)/\ (\ln \mu_d - \ln \mu_a) \tag{2.2}$$

There are many curves developed to give the value of $k$ and $h$ such as Goel and Wu [7] and Kemp [6]. However, the British Standards (BS) 5703, Part 4 is used as it is more relevant especially in counted data.

The values of $k$ and $h$ as presented in BS 5703 gives the value of $k$ to be approximately that of equation 2.2.

## 1.0 Design of the cholera outbreak CUSUM scheme in Nigeria

Ten years (1996 – 2005) reported cholera cases data collected from the WHO weekly epidemiological record is used. The schemes assume a direct proportionality between reported cases and outbreaks. That is, the greater the magnitude of the cases reported the greater the chance of an outbreak.

Two CUSUM plots for this data are constructed. These plots show the cumulative number of reported cholera cases against the cumulative number of years. For convenience the total cases reported in a year is further reduced to the number of cases per day. Between 1996 and 1998 a total of 17,160 cases were reported. That means an average of 17 per day. Between 1999 and 2001 a total of 37,920 cases were reported increasing to an average of 35 cases per day. The last four years (2002 – 2005) has a total of 15,025 cases reducing to 14 per day. For details see Appendix B.

The CUSUM scheme for the detection of the increase/decrease in the rate of cholera cases in Nigeria is therefore given by for the increasing scheme, and $S_i = \max(0, k - Y_i + S_{i-1}) \geq h$ for the decreasing scheme.

By using the reported cholera cases between 1996 and 2005, we easily detect an outbreak of cholera. From the data given in the Appendix B, the mean level of occurrence of the disease for the ten years (1996 – 2005) is taken as the mean level of the disease. The acceptable mean level $\mu_a$ is then chosen nearer to the current mean level - see Lucas (1985) [8]. From the data, the mean and the standard deviation are both 19.6. Therefore, in line with the suggestion of Lucas [8] we chose an acceptable mean $\mu_a \approx 20, \sigma = 19.6$

Suppose the management chooses a shift of $\pm 1.25\sigma$ from $\mu_a$ upward/downward to be an acceptable level and beyond it a rejectable region (the values of the mean level, acceptable mean ($\mu_a$) and the level to be detected quickly ($\mu_d$) are to be determined by the management set goals which can be revised from time to time in accordance with the management's desire). With this value of $\mu_a$ then $\mu_d \approx 45$ and $k$ is chosen to be 31 from equation 2.2. The corresponding value of $h$ from Table 1 (See Appendix A) is 19. Hence, the increasing scheme with $S_i = \max(0, Y_i - k + S_{i-1}) \geq 19, k = 31$ gives the Table 3.1 below.

**Table 3.1:** CUSUM increasing scheme for cholera detention

| **S/N** | $Y_i$ | $Y_i - k$ | $S_i\left(S_0 = 0\right)$ | $S_i\left(S_0 = 10\right)$ | **Remark** |
|---|---|---|---|---|---|
| 1 | 34 | 3 | 3 | 13 | In-control |
| 2 | 7 | -24 | 0 | 0 | " |
| 3 | 10 | -21 | 0 | 0 | " |
| 4 | 72 | 41 | 41 | $41^+$ | Out of control |
| 5 | 26 | -5 | 36 | $36^+$ | Out of control |
| 6 | 6 | -25 | 11 | 11 | In-control |
| 7 | 15 | -16 | 0 | 0 | " |
| 8 | 5 | -26 | 0 | 0 | " |
| 9 | 9 | -22 | 0 | 0 | " |
| 10 | 12 | -19 | 0 | 0 | " |

Similarly, for the decreasing scheme $[\max(0, k - Y_i + S_{i-1})] \geq 9 \;\; k = 11$ we get Table 3.2 below

**Table** 3.2: CUSUM decreasing scheme for cholera detention

| **S/N** | $Y_i$ | $Y_i - k$ | $S_i\left(S_0 = 0\right)$ | $S_i\left(S_0 = 10\right)$ | **Remark** |
|---|---|---|---|---|---|
| 1 | 34 | -23 | 0 | 0 | In-control |
| 2 | 7 | 4 | 4 | 4 | " |
| 3 | 10 | 1 | 5 | 5 | " |
| 4 | 72 | -61 | 0 | 0 | " |
| 5 | 26 | -15 | 0 | 0 | " |
| 6 | 6 | 5 | 5 | 5 | " |
| 7 | 15 | -4 | 1 | 1 | " |
| 8 | 5 | 6 | 7 | 7 | " |
| 9 | 9 | 2 | $9^*$ | $9^*$ | Out-of-control |
| 10 | 12 | -1 | 8 | 8 | In-control |

## 2.0 Discussion of findings

The cholera outbreaks (increasing scheme) will give an out-of-control signal whenever $\max(0, Y_i - 31 + S_{i-1}) \geq 19$ with ARL of 67 when the process is operating at an acceptable level and 4 when the system is unacceptable. Furthermore, the designed defection of cholera outbreaks decreasing scheme will signal out-of-control whenever $\max(0, 11 - Y_i + S_{i-1}) \geq 9$ with ARL of 24 when the process is in control and an ARL of 1.4 when the process is out of control i.e. when the system is operating at a level that should be detected quickly. The designed increasing scheme when applied to the WHO epidemiological records between 1996 and 2005 data gave an out-of-control signal between 1999 and 2000 (indicating that there were outbreaks of cholera in 1999 and 2000). Lawoyin et al [9] and many other authors confirm these outbreaks. Since then, there have been sporadic outbreaks especially in the northern part of theNigeria.

The decreasing scheme signals an excessive decrease in the year 2004. Generally, there was a significant decrease in cholera outbreaks in the last three years under review (2003 – 2005). This decrease could be as result of awareness, intensified environmental sanitation exercises encouraged by some concerned some government/governmental agencies in the various states. Prior to this time, the military enforced it with grim determination. This decline in environmental sanitation in 1999/2000 might be a significant factor for the outbreaks of cholera in those years as signaled by the designed scheme.

## 3.0 Conclusion

The designed scheme, if efficiently used, will always give an out-of-control signal whenever cholera outbreak is threatening. Under such situation, the source of drinking water should be checked; street vended water should be strictly put on check. Environmental sanitation exercise and washing of hands with soap and water should be enforced. The fore-going factors, if given attention would go a long way to reducing cholera outbreaks in Nigeria.

## APPENDIX A

**Table I** (Extracts from British Standards 5703 part 4)

| **Event rate at AQL, $m_a$** | **CUSUM Parameters for $C_1$ schemes** | | **CUSUM Parameters for $C_2$ schemes** | |
|---|---|---|---|---|
| | **H** | **K** | **H** | **K** |
| 0.1 | 1.5 | 0.75 | 2 | 0.25 |
| 0.125 | 2.5 | 0.5 | 2.5 | 0.25 |
| 0.16 | 3 | 0.5 | 2 | 0.5 |
| 0.2 | 3.5 | 0.5 | 2.5 | 0.5 |
| 0.25 | 4 | 0.5 | 3 | 0.5 |
| 0.32 | 3 | 1 | 4 | 0.5 |
| 0.4 | 2.5 | 1.5 | 3 | 1 |
| 0.55 | 3 | 1.5 | 2 | 1.5 |
| 0.64 | 3.5 | 1.5 | 2 | 2 |

| Event rate at AQL, $m_a$ | CUSUM Parameters for $C_1$ schemes | | CUSUM Parameters for $C_2$ schemes | |
|---|---|---|---|---|
| | H | K | H | K |
| 0.8 | 5 | 1.5 | 3.5 | 1.5 |
| 1 | 5 | 2 | 5 | 1.5 |
| 1.25 | 4 | 3 | 5 | 2 |
| 1.6 | 5 | 3 | 4 | 3 |
| 2 | 7 | 3 | 5 | 3 |
| 22.5 | 7 | 4 | 5 | 4 |
| 3.2 | 7 | 5 | 5 | 5 |
| 4 | 8 | 6 | 6 | 6 |
| 20 | 20 | 23 | 14 | 23 |
| 25 | 24 | 28 | 17 | 28 |
| 28 | 27 | 31 | 19 | 31 |

*$C_1$ = scheme having high ARL at AQL; $C_2$ = schemes having low ARL at AQL

## APPENDIX B:

**Table 2**: Reported cases of cholera outbreaks between 1996 and 2005.

| Year | Reported cases | Reported case per day | Average per day in the grouped years |
|---|---|---|---|
| 1996 | 12374 | 34 | 17 |
| 1997 | 1322 | 7 | |
| 1998 | 3464 | 10 | |
| 1999 | 26358 | 72 | 35 |
| 2000 | 9363 | 26 | |
| 2001 | 2199 | 6 | |
| 2002 | 5429 | 15 | 14 |
| 2003 | 1933 | 5 | |
| 2004 | 3186 | 9 | |
| 2005 | 4477 | 12 | |

### *References*


[1] Page, E. E (1954). Continuous inspection schemes, Biometrika, 41, 100 – 108.

[2] Adeyemi, R. A (2007). A Review of Statistical Models for early detection of outbreak of Disease epidemics and application, Journal of the Nig. Stat. Ass. , 19, 80 – 89.

[3] Lucas, J. M and Crossier R. B (1982). Fast Initial Response of CUSUM Quality control schemes, Technometrics, 24, 199 - 205.

[4] Lucas J. M. (1976). The Design and use of cumulative sum control schemes, Technometrics, 14, 51 – 59.

[5] Osanaiye P. A. and Edokpa I. W (1998). Control of Time – Between – Road Accidents in Kwara State of Nigeria using CUSUM Chart Techniques. Journal of Nigerian Statisticians, 11, 7 – 13.

[6] Kemp, K. W (1962).The use of cumulative sum for sampling inspection schemes, Applied Statistics, 11, 16 – 31.

[7] Geol, S and Wu (1971). Determination of ARL and a contour nomograph for CUSUM charts to control normal means, Tecnometrics 13, 221 – 230.

[8] Lucas, J. M. (1985) Counted Data CUSUM, technometrics, 27, 129 - 136.

[9] Lawoyin T. O., Oguntodede N. A., Olumide E. A and Ohadeko M. O (1999): Outbreak of Cholera in Ibadan, Nigeria. European Journal of Epidemiology, 15: 367 – 370.