A Power Analysis of Two Surveillance Methods In Terms of Average Run Lengths

Gerald Shoultz, Department of Statistics, Grand Valley State University
Paul Stephenson, Department of Statistics, Grand Valley State University
J. Wanzer Drane, Department of Epidemiology and Biostatistics, University of South Carolina

Corresponding Author: Gerald R. Shoultz Jr., Department of Statistics, Grand Valley State University, 1 Campus Drive, Allendale, MI 49401-9403, (616) 331-8689, shoultzg@gvsu.edu

ABSTRACT

This paper compares two methods for testing hypothesis usable in disease surveillance and process control to determine which is more powerful: TEXAS (Hardy et al 1980) and CUSUM (Hawkins and Olwell 1998). TEXAS, a modification of the procedures of Shewhart (1931), uses a two-step decision rule to determine when a process is under control. CUSUM finds a process to be out-of-control when the sum of a set of measurements exceeds a given threshold. First, the authors will discuss how these process control procedures can be used to monitor disease surveillance. Then the authors will present a simulation that compares the performance of the TEXAS and CUSUM methods to examine which method is more powerful for a variety of hypotheses.

INTRODUCTION

In determining whether incidence rates for a disease over time indicate a need for further in-depth study or action two concerns must be balanced. Accepting too high an incidence rate before taking further action can lead to needless deaths or suffering, with potentially high economic and emotional costs. On the other hand, reacting to relatively miniscule increases in incidence can result in needless economic costs and destruction of livelihood.

Incidence rates over time can be treated like rates or counts of defective products in a production line. Therefore, one can legitimately apply industrial quality control procedures to disease surveillance. Shewhart's methods (1931) are probably the most famous and are used frequently in industry to monitor a variety of process characteristics. Suppose w denotes a statistic of interest with mean μ_w and standard deviation σ_w. Assuming that these parameter values are known, a Shewhart-type control chart would signal that the process is out of control if a subgroup yields a computed statistic w beyond the control limits $\mu_w \pm 3\sigma_w$. Since this paper addresses incidence counts, we will focus on the c-chart, a Shewhart chart used for counts of defects in a unit of production. See Ryan (2000) for further discussion of c-charts.

Hawkins and Olwell (1998, p. 8ff) note that Shewhart charts are useful for flagging isolated, large shifts in a data set, but they are not as useful in detecting moderate, persistent shifts over time. This is because Shewhart charts are 'memoryless'—they only look at one sample at a time instead of a series of samples. As a result, numerous supplementary runs rules have been developed that lead to a signal if set of successive points have an unusual pattern. One such rule signals that the process is out-of-control if two out of three successive points are beyond 2 standard deviations from the center line.

The "TEXAS method" of Hardy et al. (1990) is a c-chart procedure with a supplementary runs rule. Hardy et al. proposed this method for "monitoring the health status of a community potentially exposed to a hazardous environment." (p. S32). An "action" level is reached when the number of cases for a time period exceeds a particularly high preset value. An "alert" level is reached when said number of cases is above a lower level (but below the "action" level). Two consecutive "alert" levels is equivalent to reaching the "action" level. The process is considered out-of-control if (1) an action level is reached in one time interval or (2) an alert level is reached in two consecutive time intervals. The TEXAS method observes incidence over a series of time intervals and uses two consecutive measures to determine whether further study or action is warranted.

Hawkins and Olwell (1998, p. 9ff) note that supplementary runs rules do not perform as well as other methods "better suited to the detection of persistent shifts in the stream of process data". A Cumulative Sum (CUSUM) procedure considers a process out of control when the accumulated number of cases over an unbroken time period exceeds a certain amount. CUSUM methods were soon recognized as more useful for detecting smaller, persistent shifts.

In this paper we describe and execute a procedure for determining whether the TEXAS or CUSUM procedure has greater statistical power (i.e., has lower Type II error) under various null and alternative hypothesis. First, as mortality and incidence counts are seen as Poisson distributed specific descriptions of TEXAS and CUSUM for Poisson-distributed data will be presented. Second, a method for equating Type I errors for TEXAS and CUSUM is presented. Such is necessary because TEXAS and other Shewhart methods state Type I errors in terms of probabilities α while CUSUM states Type I errors in terms of Average Run Length (ARL). In our context ARL is the average number of time periods one would expect to observe before incidence counts reached the action level. Third, a simulation procedure for measuring Type II errors for TEXAS and CUSUM is presented and executed. Fourth, the simulation results are presented and analyzed.

HYPOTHESES AND DEFINITIONS

Suppose that when the process is in control observations are independently and identically distributed as Poisson random variables with mean/parameter λ_1. Define the null and alternative hypotheses respectively as $H_o : \lambda = \lambda_1$ and $H_A : \lambda > \lambda_1$. Let X_i be the number of cases observed at time *i*, with $X_i \sim Poisson(\lambda_1)$. We reject H_o when a signal occurs indicating that the process is out of control.

For TEXAS define the Type I error α as the probability that the process is out-of-control in either of two consecutive measurements. Define p_{action} and p_{alert} as the probability that X_i reaches the action or alert level in a *single* time interval, respectively. In order for the alert and action levels to be different, $0 < p_{action} < p_{alert} < \alpha$. If $P(Y \geq X_i \mid Y \sim Poisson(\lambda_1)) \leq p_{action}$ then X_i has reached the action level and there is an indication that the process is out of control at time *i*. If $P(Y \geq X_i \mid Y \sim Poisson(\lambda_1)) \leq p_{alert}$ and $P(Y \geq X_{i+1} \mid Y \sim Poisson(\lambda_1)) \leq p_{alert}$ then the count of cases has reached the "alert" level at two consecutive times and there is an indication that the process is out of control at time *i*+1. One can show (Appendix A) that $p_{alert}^2 = p_{action}^2 - 2p_{action} + \alpha$,

$0 < p_{action} < \alpha/2$, and $\sqrt{\alpha} < p_{alert} < \alpha/2$. Usually α and p_{action} are pre-set (with the latter a small fractional part of α) and p_{alert} is deduced.

Turning to the CUSUM method the CUSUM C_i at time *i* is defined (Hawkins and Olwell, p. 145-6) as $C_i = \max(0, C_{i-1} + X_i - k), C_0 = 0, k = (\lambda_2 - \lambda_1)/(\ln(\lambda_2) - \ln(\lambda_1))$. We reject the null hypothesis if $C_i \geq h$, where *h* is a critical value determined by λ_1, λ_2, and a desired in-control Average Run Length (ARL). These *h*'s must be calculated numerically. In this paper we use a modified version of the ANYGETH program developed by Hawkins and Olwell (1998). Note that as X_i must be an integer the potential values of C_i, and hence the potential ARL's, are also limited (see Hawkins and Olwell, p. 107-120 for more details). To address this problem we obtain a rational number close to *k* and then use some "trial and error" to find values of k and of h that result in an ARL as close to our desired ARL as possible.

In our descriptions we use α's for TEXAS and ARL's for CUSUM. To compare the methods we must find a relationship between an α and an ARL. It is to this relationship that we now turn.

THE ARL vs. ALPHA PROBLEM

At each time *i* the status of the disease surveillance process is one of three states: no action/alert, alert, and action. The process at a given time will be in only one of these states. Hence TEXAS is a Markov Process with transition matrix (Hardy et al, p. S41) below:

	Second Period		
First Period	No Action/Alert	Alert	Action
No Action/Alert	$1 - p_{alert} - p_{action}$	p_{alert}	p_{action}
Alert	$1 - p_{alert} - p_{action}$	0	$p_{alert} + p_{action}$
Action	0	0	1

The above matrix can be put in the form

$$\begin{bmatrix} \begin{bmatrix} 1 - p_{alert} - p_{action} & p_{alert} \\ 1 - p_{alert} - p_{action} & 0 \end{bmatrix} & \begin{bmatrix} p_{action} \\ p_{alert} + p_{action} \end{bmatrix} \\ \begin{bmatrix} 0 & 0 \end{bmatrix} & \begin{bmatrix} 1 \end{bmatrix} \end{bmatrix} = \begin{bmatrix} Q_{2x2} & R_{2x1} \\ 0_{1x2} & P_{1x1} \end{bmatrix}$$

where Q_{2x2} is a matrix representing the transition probabilities among the non-terminating states (no action or alert), R_{2x1} is a matrix representing the transition probabilities for non-terminating to terminating states (action), and P_{1x1} is the probability of termination once reaching action status (i.e., 1). It follows (see Luenberger, 1979, p. 224-245 or similar text) that the Average Run Length of the process is the first element of the vector $(I_{2x2} - Q_{2x2})^{-1}\begin{bmatrix}1\\1\end{bmatrix} = \begin{bmatrix}1 + p_{alert}\\1\end{bmatrix}\frac{1}{p_{alert}^2 + p_{alert}p_{action} + p_{action}}$, easily calculated via matrix algebra methods. Hence, for a given p_{action} and p_{alert} the corresponding ARL for TEXAS is $\frac{1+p_{alert}}{p_{alert}^2+p_{alert}p_{action}+p_{action}}$.

We now equate the ARL's for CUSUM with the α's for TEXAS via the equation $ARL_{CUSUM} = \frac{1+p_{alert}}{p_{alert}^2+p_{alert}p_{action}+p_{action}}$. Note that not every α for TEXAS can be equated with every possible ARL_{CUSUM}. Let p_{action} = gα. Since $0 < p_{action} < \alpha/2$, it follows that $0 < g < 1/2$. One will find (difficult analytically but easily seen by setting up a spreadsheet and taking various values of g and α) that ARL_{CUSUM} strictly increases as *g* increases. Since $0 < p_{action} < \alpha/2$ and $\sqrt{\alpha} < p_{alert} < \alpha/2$ (earlier), $\min\{ARL_{CUSUM}\}$ occurs when $p_{action} = 0$ and $p_{alert} = \sqrt{\alpha}$, $\max\{ARL_{CUSUM}\}$ occurs when $p_{action} = p_{alert} = \alpha/2$ and $\frac{1+\sqrt{\alpha}}{\alpha} < ARL_{CUSUM} < \frac{\alpha+2}{\alpha^2+\alpha}$. Rather than choosing the ARL_{CUSUM}'s and α's and obtaining the g's we choose the ARL_{CUSUM}'s and g's and obtaining the α's. This can easily be done via Newton's Method or other procedures.

PROCEDURE

Define $H_o : \lambda = \lambda_1$ and $H_A : \lambda = \lambda_1 + 2^n\sqrt{\lambda_1}$ where $\lambda_1 \in \{0.1, 0.2, 0.5, 1, 2, 3, 5, 10, 25, 50, 100\}$, $n \in \{-3, -2, -1, 0, 1, 2, 3\}$
For the Type I errors the hypothetical ARL's to be considered for CUSUM are

$$ARL_{CUSUM\_HYPOTHETICAL} \in \left\{ \begin{matrix} 10, 20, 30, \ldots, 100, 120, 140, 160, \ldots, 300, \\ 330, 360, 390, \ldots, 600, 640, 680, 720, \ldots, 1000 \end{matrix} \right\}$$

First, for each $(\lambda_1, n, ARL_{CUSUM\_HYPOTHETICAL})$ the program ANYGETH was used to obtain exact in-control ARL_{CUSUM_ACTUAL}'s and their corresponding rational h's and k's as close as possible to the hypothetical ARL's. For each of these ARL_{CUSUM_ACTUAL}'s corresponding α's for TEXAS were calculated using the expression $ARL_{CUSUM\_ACTUAL} = \frac{1+p_{alert}}{p_{alert}^2+p_{alert}p_{action}+p_{action}}$ with p_{action} = gα, $p_{alert}^2 = p_{action}^2 - 2p_{action} + \alpha$, and $g \in \{0.01, 0.02, 0.03, \ldots, 0.49, 0.4999\}$.
Note that a null and alternative hypothesis, Type I error ARL for CUSUM and Type I error α for TEXAS can be denoted by the ordered tuple $(\lambda_1, n, ARL_{CUSUM\_ACTUAL}, g)$.

In this paper we used computer simulations to determine whether TEXAS or CUSUM is statistically more powerful. The method that is statistically more powerful will have lower Type II error; this error occurs when one does not reject the null hypothesis while the alternative hypothesis is true. Given a true alternative hypothesis the method with lower Type II error will reject the null hypothesis in the shorter period of time.

For each $\left(\lambda_1, n, ARL_{CUSUM\_ACTUAL}\right)$ assume $X_i \sim Poisson\left(\lambda_1 + 2^n \sqrt{\lambda_1}\right)$, as required for determining Type II error. Five thousand simulations of the CUSUM procedure were executed and the run times before a signal indicating that the process was out of control occurred. These times were put in order from lowest to highest; call these times $C_{[i]}$, $i = 1,2,\ldots,5000$. Then, for each $\left(\lambda_1, n, ARL_{CUSUM\_ACTUAL}, g\right)$ also assume $X_i \sim Poisson\left(\lambda_1 + 2^n \sqrt{\lambda_1}\right)$. Five thousand simulations of the TEXAS procedure were executed in the same manner as CUSUM. These times were also put in order from lowest to highest; call these times $(T \mid g)_{[i]}$, $i = 1,2,\ldots,5000$.

For each $\left(\lambda_1, n, ARL_{CUSUM\_ACTUAL}, g\right)$ we will call TEXAS more powerful than CUSUM if $(T \mid g)_{[i]} \le C_{[i]}$ for at least 4500 (90%) of the $i = 1,2,\ldots,5000$. Similarly we call CUSUM more powerful if $C_{[i]} \le (T \mid g)_{[i]}$ for at least 4500 (90%) of the $i = 1,2,\ldots,5000$. If $(T \mid g)_{[i]} = C_{[i]}$ for all $i = 1,2,\ldots,5000$ then we will say that they are equally as powerful. If none of the above hold then we say that we cannot conclude whether TEXAS is more powerful than CUSUM for a given $\left(\lambda_1, n, ARL_{CUSUM\_ACTUAL}, g\right)$.

We also examined whether TEXAS or CUSUM is more powerful for a $\left(\lambda_1, n, ARL_{CUSUM\_ACTUAL}\right)$ regardless of the α (or g). If CUSUM is more powerful for any α then for each ordered time [*i*] the run time before an out-of-control signal for CUSUM will be less than the times for TEXAS for all values of g: $C_{[i]} \le \min\{(T \mid g)_{[i]} \mid g = 0.01, 0.02, \ldots, 0.49, 0.4999\}$ for at least 4500 (90%) of the $i = 1,2,\ldots,5000$. Similarly, if TEXAS is more powerful for any α then for each ordered time [*i*] the run time before an out-of-control signal for TEXAS will be less than the times for CUSUM for all values of g: $\max\{(T \mid g)_{[i]} \mid g = 0.01, 0.02, \ldots, 0.49, 0.4999\} \le C_{[i]}$ for at least 4500 (90%) of the $i = 1,2,\ldots,5000$. Otherwise, we say that it is inconclusive which procedure is more powerful.

RESULTS

We will focus our analysis on the last case described above, in which TEXAS vs. CUSUM is examined in terms of $\min\{(T \mid g)_{[i]}\}$ and $\max\{(T \mid g)_{[i]}\}$. We start with the null hypotheses under consideration. For each $H_o : \lambda = \lambda_1$ there are seven possible alternative hypothesis (corresponding to the seven possible values of n in $H_A : \lambda = \lambda_1 + 2^n \sqrt{\lambda_1}$) and 40 possible ARL_{CUSUM_ACTUAL}'s per (λ_1, n), corresponding to the 40 $ARL_{CUSUM\_HYPOTHETICAL}$. Hence there are a total of 280 potential cases per $H_o : \lambda = \lambda_1$.

Since the Poisson distribution is a discrete distribution the values X_i coming from said distribution are also discrete. It follows that the CUSUM $C_i = \max(0, C_{i-1} + X_i - k)$ is a discrete function. The end result (see Hawkins and Olwell, p. 108-110) is that the set of possible ARL_{CUSUM_ACTUAL}'s is limited and, furthermore, that two or more $ARL_{CUSUM\_HYPOTHETICAL}$'s will have the same ARL_{CUSUM_ACTUAL}. Therefore, for our analysis duplicate $(\lambda_1, n, ARL_{CUSUM\_ACTUAL})$'s are eliminated.

Table 1 (below) summarizes the number of unique ARL_{CUSUM_ACTUAL}'s per null hypothesis in which TEXAS was more powerful, in which CUSUM was more powerful, and in which the results were inconclusive. While the vast number of cases were inconclusive TEXAS tended to be more powerful more often with smaller values of λ while CUSUM was more powerful more often for larger λ.

Table 1: Number of Cases Where Procedure Is More Powerful, By Null Hypothesis

	Result		
Null Hypothesis	TEXAS More Powerful	CUSUM More Powerful	Results Inconclusive
$\lambda = 0.1$	41	1	121
$\lambda = 0.2$	20	2	156
$\lambda = 0.3$	15	2	173
$\lambda = 0.5$	4	3	164
$\lambda = 1$	12	6	183
$\lambda = 2$	9	8	179
$\lambda = 5$	2	8	173
$\lambda = 10$	4	18	185
$\lambda = 25$	9	23	181
$\lambda = 50$	1	15	202
$\lambda = 100$	1	16	182

Turning to the exponents n in the alternative hypothesis, for each n there are 11 values of λ_1 and 40 possible ARL_{CUSUM_ACTUAL}'s per (λ_1, n), for a total of 440 potential cases per n. Again, duplicate ARL_{CUSUM_ACTUAL}'s were discarded. Table 2 (below) summarizes these results. As n went from -3 to 3 the number of cases where TEXAS was more powerful decreased until n = 2. At that point CUSUM had more cases where it was more powerful, and at n = 3 it had more cases than even the inconclusive group.

Table 2: Number of Cases Where Procedure Is More Powerful, By Value of n in the Alternative Hypothesis $H_A : \lambda = \lambda_1 + 2^n \sqrt{\lambda_1}$

	Result		
Alternative Hypothesis	TEXAS More Powerful	CUSUM More Powerful	Results Inconclusive
n = -3	48	0	362
n = -2	33	0	379
n = -1	13	0	360
n = 0	7	0	349
n = 1	2	0	276
n = 2	7	17	129
n = 3	8	85	44

DISCUSSION

In this paper we compared two methods for disease surveillance. TEXAS uses a supplementary runs rule, while CUSUM takes the cumulative sum of cases over time. The results of Table 2, in which TEXAS is more powerful for smaller, immediate shifts while CUSUM is more powerful for persistent shifts, is consistent with the comments of Hawkins and Olwell (1998) and others.

There are numerous areas for further research in terms of both quality control and simulation design. Because the Poisson distribution is discrete it is not always possible to obtain a one-to-one correspondence between a CUSUM ARL and a TEXAS α. Further examination of specific α's and/or g's is an area of further research. Replication of this study for other distributions, especially those more suitable for modeling rarer diseases (e.g., zero-inflated Poisson or negative binomial) is another potential area of further research. On another front, using different designs for the simulation is an area of further research. We performed 5000 simulations of each procedure and then ordered them to obtain results; one could also (with sufficient computer power and memory) develop a list of, for example, 10000 iterations of Poisson distribution, use the same list of iterations with TEXAS vs. CUSUM and see how many times each one gave an out-of-control signal.

BIBLIOGRAPHY

Hardy, R.J., Schroeder, G.D., Cooper, S.B., Buffler, P.A., Prichard, H.M., & Crane, M. (1990). A Surveillance System for Assessing Health Effects From Hazardous Exposures. *American Journal of Epidemiology, 132(S1),* S32-42.

Hawkins, D.M. & Olwell, D. H. (1998). *Cumulative Sum Charts and Charting for Quality Improvement.* New York: Springer-Verlag

Luenberger, D.G. (1979). *Introduction to Dynamic Systems: Theory, Models and Applications.* New York: John Wiley and Sons.

Ryan, T.P. (2000). *Statistical Methods for Quality Improvement, 2nd Edition.* New York, John Wiley and Sons.

APPENDIX A

Define $\left(NACAL\right)_i$ = no action/alert at a SINGLE time period *i*, $\left(AL\right)_i$ = alert at a SINGLE time period *i*, $\left(AC\right)_i$ = action at a SINGLE time period *i* and P(x) = the probability of an event x. Then
α = P($\left(AC\right)_1$) + P($\left(AL\right)_1$ and $\left(AL\right)_2$) + P($\left(AL\right)_1$ and $\left(AC\right)_2$) + P($\left(NACAL\right)_1$ and $\left(AC\right)_2$) = p_{action} + p_{alert} p_{alert} + p_{alert} p_{action} + $(1 - p_{alert} - p_{action})$ p_{action} = $p_{alert}^2 - p_{action}^2 + 2p_{action}$, which leads to the expression
$p_{alert}^2 = p_{action}^2 - 2p_{action} + \alpha$.

To prove $0 < p_{action} < \alpha / 2$: $0 < p_{action} < p_{alert}$ implies that $p_{action}^2 < p_{alert}^2 = p_{action}^2 - 2p_{action} + \alpha \Rightarrow 0 < -2p_{action} + \alpha \Rightarrow p_{action} < \alpha / 2.$

To prove $\sqrt{\alpha} < p_{alert} < \alpha / 2$: $p_{action} = 0 \Rightarrow p_{alert} = \sqrt{\alpha},$ $p_{action} = \alpha / 2 \Rightarrow p_{alert} = \alpha / 2,$ and procedures of elementary calculus will show that $p_{alert} = f(p_{action}) = \sqrt{p_{action}^2 - 2p_{action} + \alpha}$ is a strictly decreasing function.