International Journal of Statistical Distributions and Applications
2016; 2(4): 76-80
http://www.sciencepublishinggroup.com/j/ijsda
doi: 10.11648/j.ijsd.20160204.16
ISSN: 2472-3487 (Print); ISSN: 2472-3509 (Online)

Surveillance Methods for the Rare Health Events-A Systematic Review

Sasikumar Ramaraj, Bangusha Devi Subramanian

Department of Statistics, Manonmaniam Sundaranar University, Tirunelveli, India

Email address:
sasikumarmsu@gmail.com (S. Ramaraj), bangusha89@gmail.com (B. D. Subramanian)

To cite this article:
Sasikumar Ramaraj, Bangusha Devi Subramanian. Surveillance Methods for the Rare Health Events-A Systematic Review. *International Journal of Statistical Distributions and Applications.* Vol. 2, No. 4, 2016, pp. 76-80. doi: 10.11648/j.ijsd.20160204.16

Received: October 3, 2016; **Accepted**: November 1, 2016; **Published**: December 30, 2016

Abstract: In the field of healthscinece, the problem will arise when monitoring the incidence rate of an event is very small. There are only few statistical methods available for the investigation of rare health events. Apart from the first quality control chart introduced by Shewhart, the best well known surveillance procedures are based on the CUSUM method which was used to identify small shift in the process. But this well known surveillance procedure based on Cumulative Sum (CUSUM) method is failed to detect an increased rate when the increased rate of an event is very small. Some of the other methods like Sets method, CUSCORE (Cumulative Score) method and Bernoulli CUSUM method based were developed to carry out this problem. Detailed reviews of these three methods in the field of health science were discussed based on earlier literatures.

Keywords: Sets Method, CUSCORE Method, Bernoulli CUSUM, Steady State Average Run Length (ARL), Rare Health Event

1. Introduction

Now-a-days several fields such as industries, medicine continuous surveillance systems are required to detect an increased rate of some rare event. During last decades monitoring increased rate of rare health events is a vital curiosity in the field of medical. Several methods developed for carry out this situation and they are Sets method [5] and CUSCORE method [8] and also Bernoulli CUSUM [11]. In order to define these three methods, let us consider a sequence of Bernoulli trials V_1, V_2, V_3.... That represent each event, where V=1 indicates the presence of the particular event and V=0 indicates its absence. Let us assume that under non-increased rate the incidence rate is $p = p_0$. Under increased rate assume that the incidence rate is some unknown $p = \gamma p_0$ for suitable values of γ such that $0 < p < 1$ and $p \neq p_0$. Since we are considered here only the increased rate, let us assume that $1 < \gamma < 1/p_0$. Let $\tau \geq 0$ represent the position of the event which the shift in p takes places. Hence $P(V_j = 1) = p_0$ for $j = 1, 2, ..., \tau$ and $P(V_j = 1) = p$ for $j = \tau+1, \tau+2, ...$. The three methods considered here for detecting a shift from p_0 to $p_1 = \gamma_1 p_0$, for $1 < \gamma_1 < 1/p_0$. Here γ_1 increased rate that the chart is going to be detected. The proper descriptions of the three methods are given below with detailed literature review.

2. Sets Method

A Set is defined as the group of infants who are born consecutively between those born with a particular type of congenital malformation. Let X_i represent the size of the i[th] set, i.e. the number of births between (but not including) incidents i-1 and i for i = 1, 2,.... Then, X_1, X_2,...form a sequence of independent geometric random variables with mass function $P(X = x) = p(1-p)^x$ and take on values {0, 1,2,…}. The Sets method signals an alarm if n_s consecutive set each have a size less than a threshold t_s. It is convenient to represent the Sets method as follows. When the i[th] malformation is observed, let S_i represent the number of consecutive set that have sizes less than t_s, i.e.

$$S_0 = 0$$

$$S_i = (1 + S_{i-1}) I_{\{X_i < t_s\}}, \quad i = 1, 2, \ldots,$$

An alarm is signaled when $S_i = n_s$.

3. Cuscore Method

In the CUSCORE method, X_i, n_c, and t_c are defined as in the Sets method. This time using the c subscript to indicate the CUSCORE method. To each set size X_i, assign a score $g(X_i) = 1$ if $X_i < t_c$ and $g(X_i) = -1$ otherwise. The CUSCORE C_i is defined as

$$C_0 = 0$$

$$C_i = \max(0, C_{i-1} + g(X_i))$$

The CUSCORE method signals an alarm when $C_i = n_c$.

4. Bernoulli CUSUM Method

The Bernoulli CUSUM chart is based directly on the individual observations, X_1, X_2, … X_n without using a summary statistic based on grouping the items into samples. The Bernoulli CUSUM chart for detecting an increase in p, the Bernoulli CUSUM control statistic is

$$B_0 = 0 \quad B_j = \max(0, B_{j-1}) + (Y_j - \delta), \qquad j = 1, 2, 3, ...$$

where the reference value

$$\delta = \frac{-\log\left(\dfrac{1-p_1}{1-p_0}\right)}{\log\left(\dfrac{p_1(1-p_0)}{p_0(1-p_1)}\right)}$$

The CUSUM chart signals an alarm when Bj is greater than or equal to a control limit h.

5. Steady- State Average Run Length

Usually, the ARL denoted as the expected number of observations from the onset of monitoring until a signal is given. For the surveillance methods considered here, the observational unit is a single birth. In order to compare the Sets based methods with the Bernoulli CUSUM chart the Average Number of Births (ANB) can be used. ANB are observed from the onset of monitoring until signal. The runlength of any control chart that accumulates information depends on the starting value used for the chart statistic when monitoring begins. Naturally, if the chart starts at a value that is near the control limit, the run length will, on average, be shorter than if the chart starts at a value beyond from the control limit. If the shift occurs after a period of baseline monitoring the value of the chart statistic could be any one of its possible values when the shift occurs. A general method used in the quality control literature to account for this condition is to assume that the chart statistic has reached a stationary or steady state distribution before the shift occurs. To calculate a weighted average of all the possible ARLs that initiate from each of the possible values of the chart statistic, the steady state distribution is used. The steady-state average number of births until signal is a desirable metric for the comparisons of surveillance methods because it reflects the possible situation that a shift will happen sometime after monitoring begins, as opposed to happening prior to the beginning of monitoring or when that chart statistics are at their initial values.

6. Historical Literature

In the work (e.g., [5]) defined a method appropriate for surveillance of congenital malformation in a single hospital as well as in several hospitals. This method is based on the number of successive births occurring between the birth of a child with the particular malformation being monitored and the birth of the next child with that malformation. Such a collection of successive births is defined as a set. If each set is below a certain size then an alarm will be signaled. She compared the Sets method with the CUSUM technique. The Sets method involves on the average a slightly longer period of time for data collection of data which will initiate an alarm when there is an increase, but the overall amount of period would be smaller on the average than for the CUSUM technique. Her method was found to be comparable to the CUSUM technique for surveillance of multiple hospitals, but a little less efficient in identifying an increase in a single hospital.

Page in 1954 gave the technique consists of repeated applications of a Sequential Probability Ratio Test (SPRT). In the work (e.g. [13]) developed a modification of Page's CUSUM. They compared the system proposed by Chen [5] with their modified Page's CUSUM with. They found that in terms of the expected time elapsed between a change and its detection, the particular procedures proposed in Chen [5] take 4-32 percent more time to detect a true change, given equal false alarm rates. They base their comparison of Chen's method on the expected number of observations it takes to detect a true increase in failure rate after matching the rate of false alarms. They showed the Chen method has an inherent drawback with respect to the modified Page procedure: should H_1 be the case, the number of times after the first alarm that an alarm is not signaled is finite if one uses the modified Page procedure. For Chen's method this number is infinite. Even if this be disregarded, they pointed out that the particular procedures proposed by Chen [5] seem to do poorer than the modified Page procedure. They indicated that if one is eager to consider a Chen-type method with m = 2 or m = 1(sequence of length m, in Chen's procedure, signals an alarm) a comparison as done above was not showed the modified Page procedure to be uniformly better and also they noticed that for the alternatives considered in Chen [5], when m = 1 the modified Page procedure was uniformly worse.

In the work (e.g., [1]) the authors compared Sets method and Poisson CUSUM method through simulation technique. They applied both methods to a multi-hospital surveillance system for identifying increases in congenital malformation rates. The performance of the Sets method applied individually to hospitals and the CUSUM technique applied to the aggregate of hospitals. They determined in their study that the Poisson CUSUM has quicker signal times and greater sensivity, greater specificity and better accuracy than the Sets method.

In the work (e.g., [2]) compared Multicommunity Sets Technique (MST) and CUSUM using a computer simulation for a series of births subject to a number of congenital malformations. For that a sequence of random numbers from a uniform distribution was generated. The simulation included 6 centres. Increases in malformation frequencies were simulated for 6 out of 6 centres and then 3 out of 6 centres. MST and CUSUM were compared using computation of sensitivity, specificity and accuracy; computation of the alarm delays, namely the time intervals between the increase and the alarm and direct comparison of the performances of the techniques under testing. This study showed that MST was neither more sensitive nor more specific that CUSUM and signaled increases with greater delay. They also showed that the advantage of CUSUM for a single malformation with a very small baseline frequency and this should have been mainly appropriate for MST surveillance.

In the work (e.g., [7]) proposed an optimal procedure of the Sets method. They introduced an iterative process to find out the parameters value n and k. They matched their proposed procedure with the original Sets method and also with the Poisson CUSUM technique. They showed that the modified Sets method lead to a more efficient result over the original Sets method. They also found that when γ value is less than 4 then CUSUM method is more efficient than the modified Sets method and when γ value is higher than 4 then modified Sets method is more efficient than CUSUM method. But optimal procedure of Sets method is very difficult than the original Sets method. So in practice it is very complex to apply modified Sets method to our data.

In the work (e.g., [6]) evaluated the relative efficiency of the Sets method and CUSUM method using Monte Carlo Methods. The comparison is based on the average time delay until a true alarm, obtained by applying the two techniques to the same simulated datasets. The datasets were created to simulate the dates of diagnosis of cases in a community under raised incidence of 1.5 times the baseline level. She generated the datasets to simulate the number and the time distribution of diagnoses in a population over a 300 year period. For each baseline rate 50 data sets were generated and for each corresponding elevated rate 20 datasets were generated. These 50 datasets were evaluated for 5 annual expected number of diagnoses (AEND) 3,4,5,6 and 7. These results gave the basis for assessing the efficiency of the two methods in the detection of low-level epidemics. From the analysis, finally the author determined that the relative efficiency of the two methods depends on the baseline frequency of diagnoses and the Sets method was more efficient than the CUSUM technique when the number of cases expected in a year is less than five.

In the work (e. g.,[10]) the authors proposed a new method to monitor the rare health events. This method is similar to the assumptions of the Sets method. Proposed method and the existing method were applied to the cancer mortality data. They compared both methods for three γ values 2,4 and 6. They showed that the proposed method have much better expected time to alarm properties and also this method allowed the use of large monitoring systems consisting of implementations of the method in a number of countries of a state independently and still maintained reasonable expected times between false alarms.

In the work (e. g., [8]) developed a CUSCORE Control chart method to identify an increased rate of rare health events. This method was derived from the Sets method. Similar to Sets method this method is based on the time interval between consecutive diagnoses. By minimizing the out of control expected delay for a given tolerated rate of false alarms the involved parameters of CUSCORE method was determined. This procedure optimizes the system in terms of the expected delay until the first true alarm. In the CUSCORE method an alarm was defined in terms of the sum of scored distances between successive events of interest. The author compared the CUSCORE method with the Shewhart, Sets and CUSUM method. He compared based on the expected number of events of interest taken to detect the first true alarm after matching the rate of false alarms. The results showed that the CUSCORE method more efficient than the other competing procedures.

In the work (e. g., [19]) developed sets method to identify out-of-control situations in which the size of shift from the in-control situation is unknown. This approach is based on a simple optimality criterion which extends in a natural manner the criterion of the minimum out-of-control average run length, and also provided an accessible tool for the practitioner. An illustration to supervise the rate of occurrence of rare events was given to exponentially distributed observations.

In the work (e.g., [4]) evaluated the performance of the Sets and the CUSCORE methods when the estimate of the baseline rate is biased. This study is based on the two assumptions. 1. The baseline rate estimate is obtained during a period at which the incidence rate is stable. 2. A cluster is associated with an abrupt, rather than gradual increased incidence rate. With respect to frequency of false alarms they evaluated the effect of an underestimated baseline rate. They evaluated the effects of 5 per cent and 10 per cent bias in the estimated baseline rate for specified conditions associated with sparse data. They showed that the effect of $\pm$ 5 per cent bias in the estimate were moderated and those of 10 per cent were considerable and the effect of an overestimated baseline rate is greater on the Sets method than it is on the CUSCORE technique and the effect of an underestimated rate is greater on the CUSCORE method than it is on the Sets method. The results showed that the two methods differ also with respect to the expected time until true alarm when the specified baseline rate was unbiased and the Sets method was the more efficient in detecting a two-fold increased rate when the number of diagnoses expected annually is less than 1.62 and the CUSCORE was the more efficient method when *E*(*X*) was greater than 1.62. With an estimated baseline rate that is 5 per cent higher than the actual rate, the turning point falls from 1.62 to 1.45 when the rate was twice the baseline rate and from 5.75 to 4.34 when the rate was triples the baseline rate.

In the work(e. g., [3]) presented and evaluated the Sets method for detecting clusters or increases in incidence rates and also the Sets method was modified to permit evaluation of an animal disease-surveillance system, based on the system's detection capability as well as financial impact of the disease and its control. The author showed that the method is sensitive to the number of temporally clustered cases and the case-rate increase required to signal an alarm. This system requiring more versus fewer clustered cases needed to signal an alarm was always preferred financially for a rare disease. Finally he concluded that the financially optimal surveillance system depends on characteristics of the disease and for rare diseases, a less-sensitive detection system is financially optimal and also for more-common diseases, a more-sensitive detection system is financially optimal. He recommended that a larger number of temporally clustered cases (n) required to signal an alarm is financially preferable to a smaller number, unless a 6-fold surveillance system is being used for a common disease or a 2-fold surveillance system is being used for an expensive and common disease. He recommended that it is necessary to consider both costs of the disease and its control, probabilities of an epidemic occurring, and its expected magnitude, before selecting the number of temporally clustered cases needed to activate alarm.

In the work (e.g., [9]) proposed that a risk-adjusted version of the refined Sets method with an example was given to demonstrate its advantage over the unadjusted method. The method was illustrated with an example of surgical monitoring. They considered example data related to the performance of an experienced cardiac surgeon. The risk distribution in the surgical monitoring example has mean 0.064 and individual risks vary from 0.02 to 0.86. The data set based on 6994 operations collected from a U. K. centre for cardiac surgery over the 7 year period 1992–1998. The data consist of information on each patient including data of operation, surgeon, type of procedure and the variables that make up the pre-operative Parsonnet score. The response was taken to be the 30-day post-operative mortality rate. The result showed that the risk-adjusted chart was infact in control, whereas the unadjusted chart mistakingly signals in response to variation due to patient mix, rather than variation due to a change in the surgeon’s performance.

In the work(e. g.,[15]) compared the performance of the Sets method and its modifications with the Bernoulli CUSUM chart based on steady state ARL performance. Firstly Sego et al were used Bernoulli CUSUM chart for the surveillance of rare health event. According to satisfy a minimax criteria, chart design parameters were chosen. When the charts were designed for detecting a relatively large shift and a small one actually occurred, in which case the CUSCORE showed only slightly better performance than the Bernoulli CUSUM in a handful instances. Otherwise Bernoulli CUSUM chart was almost uniformly better than the Sets method and its modifications such as CUSCORE and SHDA method and if the charts were designed for a small shift and a large one actually occurs, the Bernoulli CUSUM and the CUSCORE substantially outperform the Sets method and SHDA method. Also they showed that all the methods were very similar when the charts were optimized for detecting large shifts.

7. Discussions

Several statistical techniques have been developed to detect epidemics of disease or birth defect in humans. The most commonly used techniques are the Sets, CUSCORE, and CUSUM method based on Binomial, Poisson and Bernoulli distributions. Many attempts have been made to prove the efficiency for the best method. From the context of our discussions, the surveillance methods ranked in the terms of the efficiency with which they accumulate information and react to a shift in the incidence rate. Particularly the Bernoulli CUCUM method is best, followed by the CUSCORE, then the Sets Method. After Sego et al. [15] argument, the Bernoulli CUSUM is considered as the convenient one for detecting increased rate of rare event. According to the argument, recently the Bernoulli CUSUM technique has been widely used by the medical practioners for detecting small shift of the rare health event. In future, if comparison will make for a prospective monitoring procedure using risk-adjustment for the Sets method, CUSCORE method and Bernoulli CUSUM, it may be more efficient than the non-risk adjustment procedures. But it may be difficult to apply in some practical situations in the field of health sciences.

8. Conclusion

Disease surveillance is an important work for health science researchers. Because it allows the researchers to detect the presence of the occurrence of unusual increase of the disease. Several methods are available to detect such an increase of the disease. Among the various methods, process control techniques have an important role to detect increase rate of the disease. In this paper, we reviewed some process control scheme that are Sets, CUSCORE and Bernoulli CUSUM methods which are used for monitoring rare event. The historical development of the Sets method and the comparisons of the Sets method with other statistical methods were also focused in this paper.

References

[1] G. Barbujani and E. Calzolari, “Comparison of two statistical techniques for the surveillance of birth defects through a Monte Carlo simulation”, Statistics in Medicine, 1984, 3, pp. 239-246.

[2] G. Barbujani, I. Ceccherini and A. Russo, “Surveillance of birth Defects: The multicommunity sets technique tested by computer simulation”, European Journal of Epidemiology, 1986, 2, pp. 52-62.

[3] T. E. Carpenter, “Financial considerations of the sets technique in animal-disease surveillance”, Preventive Veterinary Medicine, 2001, 48, pp. 155-165.

[4] R. Chen, R. R. Connelly and N. Mantel, “The efficiency of the sets and the CUSCORE techniques under biased baseline rates”, Statistics in Medicine, 1997, 16, pp. 1401- 1411.

[5] R. Chen, “A surveillance system for congenital malformations”, Journal of the American Statistical Association, 1978, 73, pp. 323-327.

[6] R. Chen, “The relative efficiency of the sets and the CUSUM techniques in monitoring the occurrence of a rare event”, Statistics in Medicine, 1987, 6, pp. 517-525.

[7] G. Gallus, C. Mandelli, M. Marchi, G. Radaelli, “On surveillance methods for congenital malformations”, Statistics in Medicine, 1986, 5, pp. 565-571.

[8] G. Radaelli, “Using the Cuscore technique in the surveillance of rare health events”, Journal of Applied Statistics, 1992, 19, pp. 75-81.

[9] O. A. Grigg and V. T. Farewell, “A risk-adjusted Sets method for monitoring adverse medical outcomes”, Statistics in Medicine, 2004, 23, pp. 1593-1602.

[10] R. R. Sitter, L. P. Hanrahan, D. Demets and H. A. Anderson, “A monitoring system to detect increased rates of cancer incidence, American journal of Epidemiology, 1990, 132, pp. 123-130.

[11] M. R. Reynolds and Z. G. Stoumbos, “A CUSUM chart for monitoring a proportion when inspecting continuously”, Journal of Quality Technology, 1999, 31, pp. 87-108.

[12] M. R. Reynolds and Z. G. Stoumbos, “A general approach to modeling CUSUM charts for a proportion, IIE Transactions, 2000, 32, pp. 515-535.

[13] R. Kenett and M. Pollak, “On sequential detection of a shift in the probability of a rare Event”, Journal of theAmerican Statistical Association, 1983, 78, pp. 389-395.

[14] S. H. Steiner, R. J. Cook and V. T. Farewell and T. Treasure, “Monitoring surgical performance using risk-adjusted cumulative sum charts”, Biostatistics, 2000, 1, pp. 441-452.

[15] L. H. Sego, W. H. Woodall and Jr. M. R. Reynold, “A comparison of surveillance methods for small incidence rates”, Statistics in Medicine, 2008, 27, pp. 1225-1247.

[16] R. Chen, “Revised values for the parameters of the sets technique for monitoring the incidence rate of a rare disease”, Methods of Information in Medicine, 1986, 25, pp. 47-49.

[17] A. G. Munford, “Control chart based on cumulative Scores”, Applied Statistics, 1980, 29, pp. 252-258.

[18] W. H. Woodall, “The use of control charts in healthcare and public health surveillance”, Journal of Quality Technology, 2006, 38, pp. 88-103.

[19] G. Radaelli, “Detection of an unknown increase in the rate of a rare event”, journal of Applied Statistics, 1996, 23, pp. 105-113.