Monitoring Poisson Count Data with Probability Control Limits when Sample Sizes are Time Varying

Xiaobei Shen,[1] Changliang Zou,[2] Wei Jiang,[3] Fugee Tsung[1]

[1]*Department of Industrial Engineering and Logistics Management, Hong Kong University of Science and Technology, Kowloon, Hong Kong*

[2]*School of Mathematical Sciences, Nankai University, Tianjin, China*

[3]*Antai College of Economics and Management, Shanghai Jiaotong University, Shanghai, China*

Received 29 August 2012; revised 3 September 2013; accepted 4 September 2013
DOI 10.1002/nav.21557
Published online in Wiley Online Library (wileyonlinelibrary.com).

Abstract: This article considers the problem of monitoring Poisson count data when sample sizes are time varying without assuming a priori knowledge of sample sizes. Traditional control charts, whose control limits are often determined before the control charts are activated, are constructed based on perfect knowledge of sample sizes. In practice, however, future sample sizes are often unknown. Making an inappropriate assumption of the distribution function could lead to unexpected performance of the control charts, for example, excessive false alarms in the early runs of the control charts, which would in turn hurt an operator's confidence in valid alarms. To overcome this problem, we propose the use of probability control limits, which are determined based on the realization of sample sizes online. The conditional probability that the charting statistic exceeds the control limit at present given that there has not been a single alarm before can be guaranteed to meet a specified false alarm rate. Simulation studies show that our proposed control chart is able to deliver satisfactory run length performance for any time-varying sample sizes. The idea presented in this article can be applied to any effective control charts such as the exponentially weighted moving average or cumulative sum chart. © 2013 Wiley Periodicals, Inc. Naval Research Logistics 00: 000–000, 2013

Keywords: average run length; exponentially weighted moving average; false alarm rate; healthcare; run length distribution; statistical process control

1. INTRODUCTION

Statistical process control (SPC) charts have been widely applied in many applications including industrial quality control, service operations management, and healthcare surveillance [20, 22]. In particular, monitoring the occurrence of a rare event from a sequence of stochastic processes has received considerable attention recently, for example, the detection of nonconformities in precise machining and manufacturing, the detection of an increase in the rate of people visiting an emergency room, the mortality rate of heart surgery [15], and the number of cancer patients [14]. In general, once the SPC charts are activated, the detection aims to issue an out-of-control (OC) signal as early as possible once an adverse event occurs. Meanwhile, the false alarm rate needs to be controlled to avoid excessive interventions of the control charts.

Correspondence to: W. Jiang (jiangwei08@gmail.com)

To detect changes in the occurrence rate of an adverse event, both the count of events recorded at regular time intervals and the corresponding sample size must be available. For example, in manufacturing quality control, a sample of products with size n_t is inspected and the number of nonconformities in the sampled products is monitored to detect possible increases in the incidence rate of nonconformities. Usually, one assumes that the count of events or nonconformities follows a (conditionally) independent Poisson distribution given the corresponding sample size. When the sample size is a constant, detecting a change in the rate could be achieved simply by detecting a change in the Poisson mean. Several control charts have been proposed including the Shewhart chart [13], the cumulative sum (CUSUM) chart [10, 21], and the exponentially weighted moving average (EWMA) chart [4, 5, 8]. In practice, these control charts have been successfully implemented in manufacturing quality control.

In some applications such as healthcare surveillance, the sample size refers to the population at risk, which, however,

often changes over time. Increasing attention has thus been paid to the problem of monitoring the occurrence rate of an adverse event with time-varying sample sizes in prospective analysis (called Phase II in SPC). Under the assumption that the time-varying sample size can be characterized by a (deterministic) logistic function, Mei et al. [12] proposed three CUSUM-based control charts taking into account the time-varying sample sizes. Shu et al. [18] compared a weighted CUSUM and conventional CUSUM procedures. Dong et al. [3] proposed to monitor the EWMA statistic of incidence rate estimator. Ryan and Woodall [17] compared CUSUM methods with the EWMA chart by Dong et al. [3] assuming that the sample size follows a uniform distribution, and suggested a modified EWMA chart by adding a lower reflecting barrier. Zhou et al. [25] proposed a new EWMA method based on weighted likelihood estimation and testing. All this work was built on the assumption that the sample size follows a prespecified random or deterministic model, which is known a priori when establishing appropriate control limits before the control charts initiate. Unfortunately, as Zhou et al. [25] pointed out, traditional control charts are very sensitive to the specification of sample sizes.

In practice, our knowledge about time-varying sample sizes is often very limited. A practical solution is to estimate the sample size distribution function based on a set of historical observations. But when the historical observations are limited, the estimation would inevitably be unreliable and model misspecification and/or estimation errors would lead to unacceptable performance of the control charts [25]. To overcome this problem, this article proposes the use of probability control limits in an EWMA control chart for monitoring Poisson count data with time-varying sample sizes in Phase II process monitoring. No matter what the (unknown) time-varying sample sizes are, the proposed EWMA chart always shares an identical run length distribution with the geometric distribution and is thus called the EWMAG chart. Essentially, the EWMAG chart uses dynamic control limits which are determined online and depend only on the current and past sample size observations. There is no need to specify any sample size models before implementation except the desired false alarm rate. Although only the EWMA-type chart is discussed in this article, the key idea can be similarly applied to any effective traditional control charts such as the CUSUM chart. The center of our proposal is to maintain the conditional probability (the probability that the charting statistic exceeds the control limit given that there is no alarm before the current time point) to the specified false alarm rate at each time point. To dynamically determine the probability control limit online, a simulation-based procedure and a Markov chain procedure are discussed.

The remainder of this article is organized as follows. We first discuss the statistical model and some previous work in Section 2. The new EWMA control chart with probability control limits is then proposed in Section 3, followed by a performance study of the proposed control chart in Section 4. Section 5 focuses on a specific healthcare surveillance example to demonstrate the application of the EWMAG chart. Finally, several remarks draw the article to its conclusion in Section 6.

2. THE STATISTICAL MODEL

Let X_t be the count of adverse events during the fixed time period $(t-1, t]$. For simplicity, we will call it the count of events at time t. Suppose X_t independently follows the Poisson distribution with the mean θn_t conditional on n_t, where θ and n_t denote the occurrence rate of the event and the sample size at time t, respectively. To detect an abrupt change in the occurrence rate from θ_0 to another unknown value $\theta_1 > \theta_0$ at some unknown time τ, we use the following change-point model,

$$X_i \overset{\text{i.d}}{\sim} \begin{cases} \text{Poisson } (\theta_0 n_i | n_i) & \text{for } i = 1, \ldots, \tau - 1 \\ \text{Poisson } (\theta_1 n_i | n_i) & \text{for } i = \tau, \ldots, \end{cases} \tag{1}$$

where the symbol $\overset{\text{i.d}}{\sim}$ denotes "independently distributed." The objective is to detect the change as soon as possible after it occurs through the sequential counts.

In the change-point detection problem, a detection rule is often characterized by a charting statistic $a(\mathbf{n}_t, \mathbf{X}_t)$ and a control limit $h(\mathbf{n}_t)$ determined based on the historical data set $\{n_i, X_i\}_{1 \le i \le t}$, where $\mathbf{n}_t = \{n_i : 1 \le i \le t\}$ and $\mathbf{X}_t = \{X_i : 1 \le i \le t\}$. The stopping time T is defined as

$$T = \min \{t : a(\mathbf{n}_t, \mathbf{X}_t) > h(\mathbf{n}_t)\}. \tag{2}$$

$T = t$ means that an alarm (OC signal) is issued at time t for the first time to declare that a change has occurred at some point during the time period $[1, t]$. Similar to the literature on healthcare surveillance, this study focuses on the detection of an increase in the occurrence rate, that is, $\theta_1 > \theta_0$. Thus only the upward shift is studied in this article. The detection of downward and two-sided shifts can be constructed similarly without much difficulty. Note that the control limit $h(\mathbf{n}_t)$ depends only on $\mathbf{n}_t$ and not $\mathbf{X}_t$.

Several control charts have been developed for Poisson count data in previous studies. In the following, we will discuss in detail two EWMA charts proposed by Dong et al. [3] and Ryan and Woodall [17], respectively, and one CUSUM chart suggested by Mei et al. [12].

The statistic of the EWMA-type control chart proposed by Dong et al. [3] is

$$Z_i = (1 - \lambda) Z_{i-1} + \lambda \frac{X_i}{n_i}, \tag{3}$$

where $i = 1, 2, \ldots, t$, $Z_0 = \theta_0$, and $\lambda \in (0, 1]$ is a smoothing parameter which determines the weights of past observations. Derived from the EWMA sequence, the first EWMA control chart proposed by Dong et al. [3], termed EWMAe, has a stopping rule,

$$T_{\text{EWMAe}} = \min\{t; Z_t \geq \theta_0 + L\sigma_t, t \geq 1\},$$
$$\sigma_t^2 = \lambda^2 \sum_{i=1}^{t} (1-\lambda)^{2t-2i} \frac{\theta_0}{n_i}. \quad (4)$$

The control limit constant L is determined by the nominal value of the in-control (IC) average run length (ARL), usually termed ARL_0. To avoid the problem of inertia, Ryan and Woodall [17] modified the EWMAe method by adding a lower reflecting barrier at $Z_t = \theta_0$, that is,

$$T_{\text{EWMAM}} = \min\left\{t; \tilde{Z}_t \geq L\sigma_t, t \geq 1\right\}, \quad (5)$$

where

$$\tilde{Z}_t = \max\left\{\theta_0, (1-\lambda)\tilde{Z}_{t-1} + \lambda\frac{X_t}{n_t}\right\}, \tilde{Z}_0 = \theta_0.$$

This method is referred to as the EWMA-modified (EWMAM) control chart. The CUSUM chart proposed by Mei et al. [12] is given by

$$W_t = \max\left\{0, W_{t-1} + \left[X_t \log\frac{\theta_1}{\theta_0} - n_t(\theta_1 - \theta_0)\right]\right\},$$

which sounds an alarm when $W_t > L_C$, where $W_0 = 0$. Its control limit L_C is determined to achieve the desired ARL_0.

To determine the control limit or control limit constant for these control charts with time-varying sample sizes, the distribution of the sample size n_t has to be assumed known a priori. As discussed in Section 1, there is usually very little foreknowledge about the future distribution of sample sizes, especially when the population at risk may be subject to sudden changes due to certain events such as an outbreak of war, famine, or natural disaster. Once the assumption on the distribution deviates significantly from reality, the control limit determined based on the assumption becomes inappropriate and may result in undesired false alarm rates accordingly. This in turn hurts an operator's confidence in valid alarms. To address this issue, an EWMA control chart with probability control limits is proposed in the next section.

3. AN EWMA CHART WITH PROBABILITY CONTROL LIMITS

3.1. The EWMAG Chart

We use the EWMA-type control chart statistic (3) as the charting statistic in the following discussion of the probability control limits. The proposed EWMA control chart with probability control limits is referred to as an EWMAG chart because its IC run length distribution is theoretically identical to the geometric distribution, that is, the false alarm rate does not depend on the time of monitoring, nor the sample sizes being monitored.

The control limit of the EWMAG chart is set so that the conditional probability, that is, the probability that the charting statistic exceeds the control limit given no prior alarms, is equal to a specified false alarm rate. To be more specific, we want to find the control limits satisfying the following equations,

$$\Pr\left(Z_1 > h_1(\alpha) \mid n_1\right) = \alpha,$$
$$\Pr\left(Z_t > h_t(\alpha) \mid Z_i < h_i(\alpha), 1 \leq i < t, n_t\right) = \alpha \quad \text{for} \quad t > 1, \quad (6)$$

where α is the prespecified false alarm rate. This is equivalent to performing a hypothesis test with the type-I error α at each time point t. Therefore, the corresponding IC run length distribution is exactly the geometric distribution [6]. At time t, the probability control limit is determined right after we observe the value of n_t. Consequently, the EWMAG chart does not need to know future sample sizes and is not concerned about wrong assumptions being made. This property makes the proposed EWMAG chart significantly different from previous control charts.

It is worth noting that the idea of using time-varying control limits has been adopted in the literature of selfstarting control schemes. As indicated in those studies, the probabilities of false alarms from a chart may increase dramatically after short-runs if a fixed control limit is applied. The approach of using dynamic control limits was originally proposed by Margavio et al. [11] and Lai [9] and has been successfully formalized and utilized by Hawkins et al. [7] in the parametric change-point-based control charts with unknown IC parameters. See also Zou and Tsung [24] for a related discussion. However, it should be emphasized here that our procedure differs from all of the above in the sense that the control limits in our procedure are determined online along with the process observations rather than decided before monitoring. That is, the control limits are data dependent. This is a unique feature of our procedure and is designed to tackle the time-varying population size n_t.

Due to the intricacy of the conditional probability (6), solving $h_t(\alpha)$ analytically is tricky. Thus two computational procedures, a simulation-based one and a Markov chain one, are suggested to approximate $h_t(\alpha)$. Though only the EWMA-type control chart is discussed here, the two computational procedures can be applied to other charts such as the CUSUM chart with probability control limits.

3.2. Computational Procedures for Probability Control Limits

First, we introduce the simulation-based procedure for computing the probability control limits. To explain this procedure clearly, let us start by considering the first time point $t=1$. Under the IC condition, X_1 should follow the Poisson distribution with mean $\theta_0 n_1$, where n_1 is known exactly. Therefore, we can obtain the control limit at the first time point by randomly generating $\hat{X}_{1,i}$, where $i = 1, \ldots, M$ and M is a sufficiently large integer, from the distribution Poisson $(\theta_0 n_1)$ and correspondingly calculating M values of pseudo Z_1 from (3) with $Z_0 = \theta_0$, say $\hat{Z}_{1,1}, \ldots, \hat{Z}_{1,M}$. We then sort those values in ascending order and store them in a vector $\hat{\mathbf{Z}}_{1M}$. The control limit $h_1(\alpha)$ can be approximated by the $M' = \lfloor M(1-\alpha) \rfloor$ largest value in $\hat{\mathbf{Z}}_{1M}$, where $\lfloor A \rfloor$ denotes the largest integer less than or equal to A. Theoretically, if $M \to \infty$, the equation $\Pr(Z_1 > h_1(\alpha)) = \alpha$ can be exactly satisfied. In this article, M is set to 50,000 to obtain an appropriate control limit at each time point. After determining the control limit $h_1(\alpha)$, we compare the value of $\hat{Z}_1$, which is calculated based on the observed X_1 and n_1, with $h_1(\alpha)$. An OC signal is issued if $\hat{Z}_1 > h_1(\alpha)$. Otherwise, we continue to the next time point $t=2$.

According to (6), to determine the control limit $h_2(\alpha)$, we should ensure that the value of pseudo Z_1 is less than or equal to h_1. Hence, only the ranked values $\hat{Z}_{1,(1)}, \ldots \hat{Z}_{1,(M')}$ should be kept to determine $h_2(\alpha)$. We store the M' ranked pseudo Z_1 into a vector $\hat{\mathbf{Z}}'_{1M'}$. Given n_2, a vector $\hat{\mathbf{Z}}_{2M}$ with the dimension M can then be obtained by

$$\hat{Z}_{t,i} = (1-\lambda)\hat{Z}_{t-1,j} + \lambda \frac{\hat{X}_{t,i}}{n_2} \tag{7}$$

where $t = 2, i = 1, \ldots, M, \hat{Z}_{1,j}$ is uniformly selected from $\hat{\mathbf{Z}}'_{1M'}$ with $j \in \{1, \ldots, M'\}$, and $\hat{X}_{2,i}$ are randomly generated from Poisson $(\theta_0 n_2)$. By sorting the M elements of $\hat{\mathbf{Z}}_{2M}$ in ascending order, we can obtain the control limit $h_2(\alpha)$ by setting it at the $(1-\alpha)$-quantile of the M elements. Again we keep the ranked statistics $\hat{Z}_{2,(1)}, \ldots, \hat{Z}_{2,(M')}$ to the next stage $t=3$. Repeat the above procedure by simulating M samples of Poisson $(\theta_0 n_3), \ldots,$ and so forth. The simulation-based procedure is summarized in the following algorithm:

Algorithm 1 (Simulation-based procedure)

1. If there is no OC signal at time $t-1$ $(t = 1, 2, \ldots), \hat{X}_{t,i}\ (i = 1, \ldots, M)$ are generated from the distribution Poisson $(\theta_0 n_t)$ where n_t is known exactly. Accordingly, M values of the pseudo charting statistic Z_t are obtained through (7).
2. Sort the M values in ascending order and the α upper empirical quantile of those M values is used for approximating the control limit $h_t(\alpha)$.
3. Compare the value of $\hat{Z}_t$, which is calculated based on the observed X_t and n_t, with $h_t(\alpha)$ to decide whether to issue an OC signal or to continue to the next time point.
4. If it is decided to continue, set $M' = \lfloor M(1-\alpha) \rfloor$ and eliminate the values $\hat{Z}_{t,(M'+1)}, \ldots, \hat{Z}_{t,(M)}$. Go back to step 1.

Next, we turn our attention to the Markov chain procedure for computing the probability control limits. The Markov chain model described here can be considered as an extension of the methods proposed by Brook and Evans [2] and Borror et al. [1]. However, different from the previous methods, our Markov chain procedure is designed specially for monitoring the occurrence rate when sample sizes are time varying. Before discussing this procedure in detail, we first introduce a critical idea of this procedure, as well as some concepts including bounds of charting statistics and states.

At time t, the value of charting statistic Z_t is within an interval with two-sided bounds. As $Z_t = (1-\lambda)Z_{t-1} + \lambda X_t/n_t$ and $X_t \geq 0$, we can simply set the lower bound $L=0$ and search for the upper bound U with the constraint $\Pr(Z_t \leq U) = 1-\xi$, where $\xi > 0$ is a sufficiently small constant at each time point. In this study, we set $\xi = 1e^{-16}$. After determining the two-sided bounds, we divide the interval (L, U) equally into K partitions (states). Then the ith state is the subinterval (L_i, U_i), where

$$L_i = L + \frac{(i-1)(U-L)}{K} \text{ and } U_i = L + \frac{i(U-L)}{K}.$$

The corresponding midpoint, m_i, of the ith state can be determined by $m_i = L + (2i-1)(U-L)/2K$. When Z_t falls into the ith state at the time t, we have $L_i < (1-\lambda)Z_{t-1} + \lambda X_t/n_t \leq U_i$, which is equivalent to

$$\frac{n_t\left[L_i - (1-\lambda)Z_{t-1}\right]}{\lambda} < X_t \leq \frac{n_t\left[U_i - (1-\lambda)Z_{t-1}\right]}{\lambda}. \tag{8}$$

The probability that Z_t is within the ith state, conditioned on $Z_{t-1} = m_j\ (i, j \in \{1, \ldots, K\})$, can be obtained by calculating the corresponding probability of X_t as its probability density function is $f(X_t) = e^{-\theta_0 n_t}(\theta_0 n_t)^{X_t}/X_t!$.

To clearly describe the procedure, we again start by considering the first time point $t=1$. As Z_0 is specified as θ_0, we consider that Z_0 is in the point θ_0 with probability 1 at time $t=0$. At time $t=1$, given Z_0 and n_1, we first determine the upper bound U with the constraint

$$\Pr(Z_1 \leq U) = \Pr\left(X_1 \leq \frac{n_1(U-(1-\lambda)Z_0)}{\lambda}\right) = 1-\xi,$$

and then divide the interval into K states. The probability that Z_1 falls into the ith state can be easily calculated by

$$\Pr\,(Z_1 \in \text{state}\, i \,|Z_0) = \begin{cases} \Pr\left(\frac{n_1[L_i-(1-\lambda)Z_0]}{\lambda} < X_1 \le \frac{n_1[U_i-(1-\lambda)Z_0]}{\lambda}\right), \\ \qquad \text{if } \lfloor\frac{n_1[L_i-(1-\lambda)Z_0]}{\lambda}\rfloor < \lfloor\frac{n_1[U_i-(1-\lambda)Z_0]}{\lambda}\rfloor, \\ 0, \qquad \text{otherwise}, \end{cases} \tag{9}$$

where $i = 1, \ldots, K$. We then obtain a vector of K probabilities at time $t = 1, \boldsymbol{P}_1 = (p_1, p_2, \ldots, p_K)'$, where the element p_i is equal to the value of $\Pr\,(Z_1 \in \text{state}\, i\,|Z_0)$. The control limit $h_1\,(\alpha)$ at time $t = 1$ is determined as the upper bound U_r of the rth state, where

$$r = \arg\min\left\{ j : \sum_{i=1}^{j} p_i \ge 1-\alpha \right\}, j = 1, 2, \ldots, K. \tag{10}$$

If $\hat{Z}_1 < h_1\,(\alpha)$ at time $t = 1$, where $\hat{Z}_1$ is calculated based on the observed X_1 and n_1, the process is declared as IC and we can proceed to the next stage $t = 2$. Otherwise, an alarm should be issued.

At time $t = 2$, we first deal with Z_1 because Z_2 is partially dependent on it. The restriction of $Z_1 < h_1\,(\alpha)$ requires us to keep only the r states of Z_1 and store their normalized probabilities $\tilde{p}_j = p_j / \sum_{j=1}^{r} p_j\ (j = 1, \ldots, r)$ into a vector $\tilde{\boldsymbol{P}}_1 = (\tilde{p}_1, \ldots, \tilde{p}_r)'$. For any value of Z_1 in the jth state, it is represented by $Z_{1j} = m_j$ where $j = 1, \ldots, r$. Then, all possible values of Z_1 considered at time $t = 2$ are the r elements of $\boldsymbol{Z}_1 = (Z_{11}, Z_{12}, \ldots, Z_{1r})'$. Given $Z_1 = Z_{1r}$ and n_2, the upper bound U of Z_2 can then be determined with the constraint

$$\Pr\,(Z_2 \le U) = \Pr\left(X_2 \le \frac{n_2\,[U - (1-\lambda)\,Z_{1r}]}{\lambda}\right) = 1 - \xi. \tag{11}$$

Again, we divide the interval $(0, U)$ into K states. When $Z_1 = Z_{1j}$, K conditional probabilities of Z_2 are obtained and stored in a vector $\boldsymbol{P}_{2j} = \left(p_{1j}, p_{2j}, \ldots, p_{Kj}\right)'$, where $p_{ij} = \Pr\left(Z_2 \in \text{state}\, i\,|Z_{1j}, n_2\right), i = 1, \ldots, K$ and $j \in \{1, \ldots, r\}$. Let p_i be the probability that Z_2 falls into the ith state given that Z_1 is IC. The vector of conditional probabilities of $Z_2, \boldsymbol{P}_2 = (p_1, p_2, \ldots, p_K)'$, is calculated as

$$\boldsymbol{P}_2 = (\boldsymbol{P}_{21}, \boldsymbol{P}_{22}, \ldots, \boldsymbol{P}_{2r})_{(K\times r)}\tilde{\boldsymbol{P}}_{1(r\times 1)}. \tag{12}$$

With the conditional probabilities of Z_2, the control limit $h_2\,(\alpha)$ at time $t = 2$ is approximated by the upper bound of the rth state satisfying (10). The procedure is summarized in the following Algorithm 2.

Algorithm 2 (Markov chain procedure)

1. If there is no OC signal at time $t-1$ ($t = 1, 2, \ldots$), calculate the vector $\tilde{\boldsymbol{P}}_{t-1}$ storing r normalized probabilities and the vector $\boldsymbol{Z}_{t-1}$ containing the values of r midpoints. Both vectors are of the size $(r \times 1)$. Specially, at time 0, we have $r = 1, \tilde{P}_0 = 1$, and $Z_0 = \theta_0$.
2. Always set the lower bound, L, to 0 and search for the upper bound, U, of Z_t based on the value of $Z_{(t-1)r}$ and divide the interval (L, U) into K states.
3. For each $Z_{(t-1)j}, j = 1, \ldots, r$, compute the probability vector $\boldsymbol{P}_{tj}$ of size $(K \times 1)$. Then the vector of conditional probabilities can be obtained through formula (12).
4. Compare the value of $\hat{Z}_t$ calculated based on the observed X_t and n_t with the determined control limit at time t. The control limit, $h_t\,(\alpha)$, is chosen to be the upper bound of the rth state satisfying (10). If $\hat{Z}_t > h_t\,(\alpha)$, an OC signal should be issued. Otherwise, go back to step 1 and continue to the next time point.

3.3. Comparison of the Two Computational Procedures

As discussed before, theoretically the EWMAG chart has an identical IC run length distribution to that of the geometric random variable, although its design does not require any knowledge of n_t. To verify this statement and compare the two computational procedures, we conduct simulation studies under various scenarios of sample size. In healthcare surveillance, Mei et al. [12] suggested modeling population growth using the logistic model. A constant sample size [3] and a uniformly distributed n_t [17] have also been considered. In particular, the following four scenarios are used in our simulation studies.

(I) Increasing Scenario: $n_t = \frac{c_1}{C(0.5+\exp\{-(t-c_2)/c_3\})}$, where $C = 1$ or 8;
(II) Decreasing Scenario: $n_t = \frac{c_1/2.4}{1+\exp\{(t-c_2)/c_3\}} + C$, where $C = 1$ or 7;
(III) Constant Scenario: $n_t = 4.5$ or $n_t = 10$;
(IV) Uniform Scenario: $n_t \sim U\,(1, 4)$ or $n_t \sim U\,(5, 18)$,

In Scenarios (I) and (II) $c_1 = 13.8065, c_2 = 11.8532$, and $c_3 = 26.4037$ which are the same as those in Mei et al. [12]. Notice that each scenario is set to have different parameters, for example, different C's and different constant values. For conciseness, we choose one setting only in each scenario for testing in this section.

Under each scenario, we set the desired false alarm rate $\alpha = 0.0027$ and accordingly the desired ARL_0 should be

Table 1. IC performance of the EWMAG chart with the simulation-based procedure; $\theta_0 = 1$.

Distribution	ARL_0	SE	SDRL	Q(.10)	Median	Q(.90)	FAR
(I) [C = 8]	372	1.66	372	40	258	857	0.0781
(II) [C = 1]	371	1.67	373	39	256	854	0.0805
(III) [$n_t = 4.5$]	370	1.65	369	39	258	849	0.0805
(IV) [U(1,4)]	369	1.66	370	37	256	846	0.0826
Geometric	370	—	370	39	256	852	0.078

approximately 370. For illustration, the control chart performance is summarized using ARL_0, percentiles of the marginal distribution of the run length, and standard deviation of the run length (SDRL). Besides these quantities, we also study the false alarm rate (FAR) for the first 30 observations, $P\,(T \leq 30|\text{in} - \text{control}\,)$. We set $\theta_0 = 1$ and the smoothing parameter $\lambda = 0.1$ and use Monte Carlo simulations of 50,000 replications to estimate the run length distribution of the EWMAG chart. The Fortran codes for implementing the EWMAG chart is available from the authors upon request.

Tables 1 and 2 summarize the simulation results of the EWMAG charts using the simulation-based and Markov chain procedures, respectively. M in the simulation-based procedure is set as 50,000 and K in the Markov chain procedure is chosen as 3000. We use the notations SE , Q (.10) , Q (.90) and FAR for the standard error of ARL_0 estimation, the 10th percentile, the 90th percentile, and the false alarm rate, respectively. The IC run length distribution is considered to be satisfactory here if it is close to the geometric distribution or more generally if it varies to a lesser extent than the geometric distribution. As a reference, when the run length distribution is geometric, the SDRL should be approximately equal to ARL_0 and Q (.10) , Q (.90), and FAR should be about 39, 852, and 0.078, respectively.

Under Scenarios (I)–(IV), the values of ARL_0 obtained from the EWMAG charts based on the two procedures are apparently close to the desired value of 370 (the slight deviation is due to simulation errors). SDRL , Q (.10) , Q (.90), median, and FAR are all approximately equal to the respective theoretical values. That is, the EWMAG chart has an identical IC run length distribution to the geometric distribution and the two proposed calculation procedures have a similar design and perform similarly well. Therefore, we will use only the simulation-based procedure in the following studies when evaluating the EWMAG chart because it is faster in implementation.

4. PERFORMANCE COMPARISON

In this section, the performance of the EWMAG chart is compared with that of the EWMAe, EWMAM, and CUSUM charts under the four scenarios of time-varying sample sizes discussed previously. We set the false alarm rate $\alpha = 0.0027$ and choose the smoothing parameter of EWMA-type control charts as $\lambda = 0.1$, and $\theta_0 = 1$ of the CUSUM chart as in Mei et al. [12]. To make a comprehensive comparison, we consider the performance under both the IC and OC situations.

4.1. IC Performance

Assume that population sizes are known exactly. Figure 1 presents the IC run length distribution of the four control charts (EWMAe, EWMAM, CUSUM, and EWMAG) under the four scenarios considered in Section 3.3. As expected, the run length distribution curve of the EWMAG chart merges with that of the geometric distribution, which verifies again that the IC run length distribution of the proposed EWMAG chart is exactly the geometric distribution. In contrast, the curves of the other three control charts deviate significantly from the geometric distribution curve to different degrees under various scenarios. In particular, the EWMAe chart often has higher probabilities of giving false alarms than the geometric distribution, especially under Scenarios (I), (III), and (IV), whereas the CUSUM chart has considerably lower probabilities of doing so than the geometric distribution under Scenario (I).

The IC ARLs of the EWMAe, EWMAM, and CUSUM charts with misspecified models are reported in Table 3.

Table 2. IC performance of the EWMAG chart with the Markov chain procedure; $\theta_0 = 1$.

Distribution	ARL_0	SE	SDRL	Q(.10)	Median	Q(.90)	FAR
(I) [C = 8]	370	1.65	369	39	258	843	0.0785
(II) [C = 1]	372	1.66	370	40	259	858	0.0765
(III) [$n_t = 4.5$]	371	1.67	373	39	257	852	0.0796
(IV) [U(1,4)]	372	1.67	373	40	257	852	0.0769
Geometric	370	—	370	39	256	852	0.078

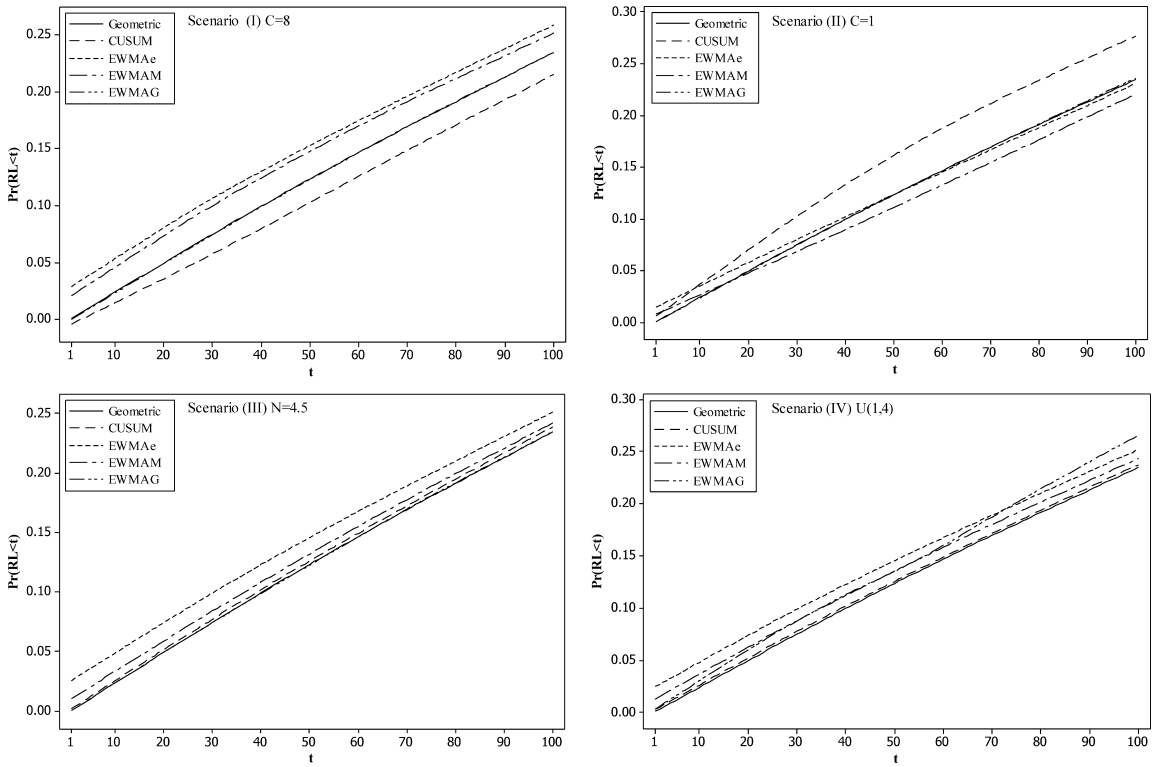


Figure 1. Comparison of the IC run length distribution among the EWMAe, EWMAM, CUSUM, and EWMAG charts.

Obviously, the observed ARL_0's would deviate from the nominal one (370) to various degrees when the distributional model of population sizes is incorrectly specified. Even with appropriate models, misspecified parameters in the distribution function also result in poor IC performance. As pointed out before, accurate information about future population sizes can rarely be obtained in many applications. Therefore, control charts constructed based on the basis that distribution functions of varying population sizes are exactly known will result in unacceptable run length distributions as shown in this table. This clearly indicates the advantage of our EWMAG chart.

Table 3. The effect of misspecified population sizes on IC ARLs of the EWMAe, EWMAM, and CUSUM charts; $\theta_0 = 1$.

	Real distribution							
Assumed Dist.	(I) [C = 8]	(I) [C = 1]	(II) [C = 1]	(II) [C = 7]	(III) [n_t = 4.5]	(III) [n_t = 10]	(IV) [U(1,4)]	(IV) [U(15,18)]
EWMAe								
(I) [C = 8]	/	461	307	407	384	424	343	437
(II) [C = 1]	450	566	/	502	471	521	411	553
(III) [n_t = 4.5]	352	436	295	390	/	408	327	426
(IV) [U(1,4)]	393	492	325	437	416	461	/	486
EWMAM								
(I) [C = 8]	/	486	294	418	392	441	331	479
(II) [C = 1]	467	648	/	549	503	580	423	615
(III) [n_t = 4.5]	347	459	286	398	/	420	319	435
(IV) [U(1,4)]	408	545	329	472	437	495	/	526
CUSUM								
(I) [C = 8]	/	1032	503	424	402	390	394	607
(II) [C = 1]	282	986	/	256	292	374	300	481
(III) [n_t = 4.5]	314	1036	499	423	/	394	390	589
(IV) [U(1,4)]	297	1019	466	419	291	393	/	561

Table 4. OC ARL comparison of the EWMAG chart and the EWMAe chart under Scenarios (I)–(IV) with $\tau = 21$.

	(I) [C = 8]		(II) [C = 1]		(III) [$n_t = 4.5$]		(IV) [U(1,4)]	
θ_1	EWMAe	EWMAG	EWMAe	EWMAG	EWMAe	EWMAG	EWMAe	EWMAG
1.025	249	241	284	285	233	224	266	259
1.050	171	168	220	217	149	146	194	188
1.075	126	124	172	169	104	99.9	146	141
1.100	95.3	94.9	137	133	73.8	72.2	111	107
1.200	44.5	43.8	55.2	52.5	27.7	27.8	48.0	45.9
1.300	27.2	27.1	26.3	25.0	15.2	15.1	26.7	25.9
1.400	19.2	19.1	15.1	14.5	10.0	9.88	17.4	17.1
1.500	14.5	14.5	10.3	9.89	7.31	7.25	12.6	12.6
1.750	8.77	8.81	5.50	5.32	4.19	4.15	7.11	7.18
2.000	6.18	6.18	3.64	3.49	2.83	2.80	4.81	4.90
2.500	3.69	3.72	2.03	1.96	1.59	1.59	2.77	2.94
3.000	2.56	2.58	1.33	1.27	1.02	1.02	1.87	2.06
4.000	1.48	1.48	0.66	0.63	0.47	0.47	1.01	1.23

4.2. OC Performance

To investigate the OC performance, only the EWMAe chart is used for the comparison because it has an identical form to the EWMAG chart except for the control limits. We assume that the population sizes are known exactly here as it is unfair to compare different procedures in terms of OC ARL when their IC run length distributions differ significantly. In general, OC run length distributions depend on the OC conditions (i.e., the rate of event occurrence changes from θ_0 to θ_1 at time τ) and the occurrence time τ. Therefore, in the following, OC performance is studied under different values of θ_1 and occurrence time τ successively.

Assuming $\tau = 21$, Table 4 presents the OC ARLs of the EWMAG and EWMAe charts with different values of θ_1 under Scenarios (I)–(IV). It is easy to see that the two control charts are comparable in terms of OC performance regardless of the population scenarios. This demonstrates that the proposed EWMAG chart can deliver the desired IC run length performance without degradation of its change detection ability.

Setting $\theta_0 = 1.2$, Fig. 2 presents the OC ARLs of the two charts under Scenarios (I)–(IV) when the occurrence time τ ranges from 1 to 100. The figure shows that both charts are sensitive to the occurrence time of the change, especially when population monotonically increases or decreases. In general, the EWMAG chart always has smaller OC ARL values than the EWMAe chart, except when τ is very small under Scenario (I).

At this point, it is worth mentioning that the main objective of the proposed dynamic procedure is to make the IC run length distribution of a control chart attaining the theoretical geometric distribution rather than to improve the detection ability of the chart. Therefore, we suggest using the proposed EWMAG chart due to its desired IC run length performance and competitive OC performance.

5. AN EXAMPLE IN HEALTHCARE SURVEILLANCE

In this section, an example of female thyroid cancer in healthcare surveillance is used to demonstrate the application of the proposed EWMAG chart. According to a report from the National Cancer Institute, there are about 37,000 new cases of thyroid cancer each year in the United States and females are most likely to have thyroid cancer at a ratio of 3:1. Thyroid cancer may occur in any age group, although it is most common after age 30, and its aggressiveness increases significantly in older patients.

The data, provided by the New York State Cancer Registry through its official website[1], include the number of female thyroid cancer cases and the incidence rate each year in the New York State. Based on the provided data, the corresponding population size each year can be easily derived. In Fig. 3, (a) and (b) show the time series plots of the counts (in the units of 100 cases) and the incidence rates per 100 million population of female thyroid cancer in the New York State, respectively. It can be observed that the incidence rate was quite stable before 1982 but exhibited a slight increase after 1983. From 1990 onward, the increase became more significant. In Fig. 3c, the female population significantly increased from 8.95 million in 1976 to 9.88 million in 1995.

Based on the pattern of incidence rate discussed above, the period from 1976 to 1982 is chosen as the Phase-I reference sample and accordingly the nominal incidence rate is estimated as $\theta_0 = 0.45$. A calibration sample of this size may not be large enough to precisely determine the IC parameter, but it suffices to illustrate the use of the method in a real-world setting. Our aim is to monitor the incidence rate of female

[1] http://www.health.ny.gov/statistics/cancer/registry/table2/tb2thyroidnys.htm.

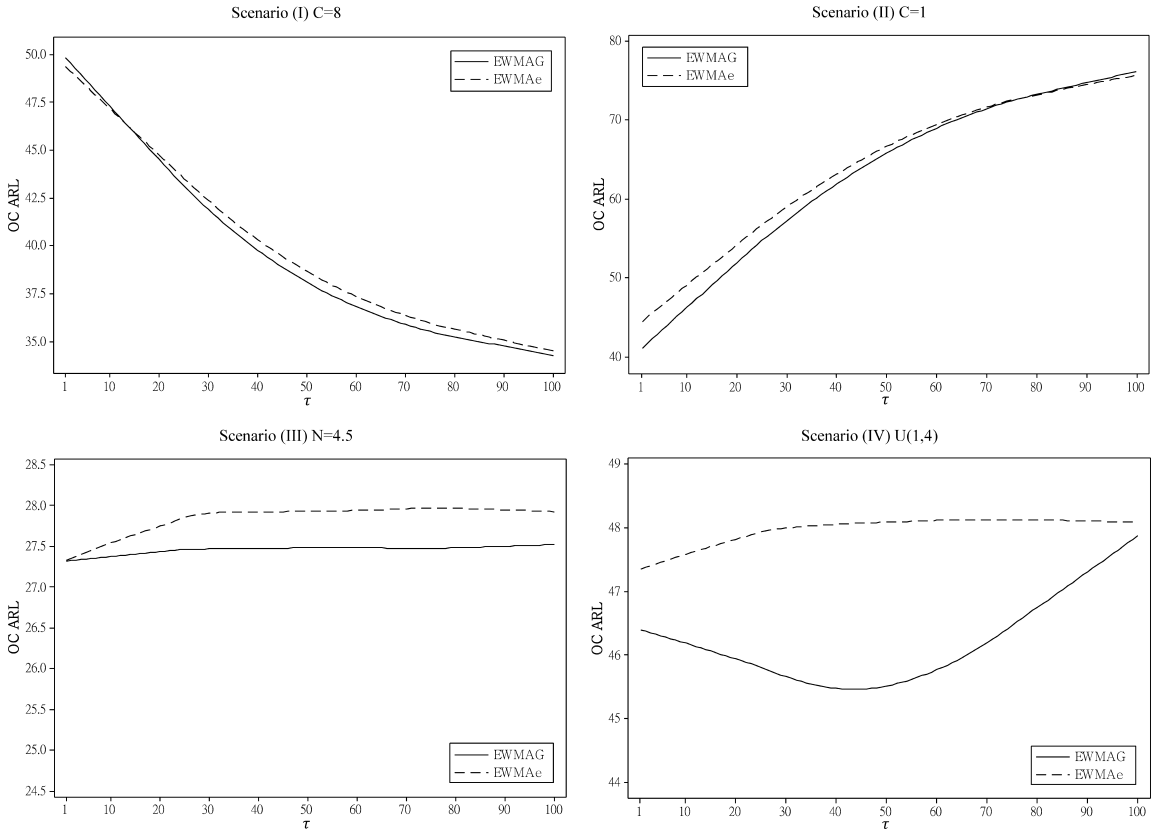


Figure 2. OC comparison of the EWMAe and EWMAG charts with different τ.

thyroid cancer from 1983 to 1995 and compare the performance of the EWMAG chart and that of the EWMAe chart in this example.

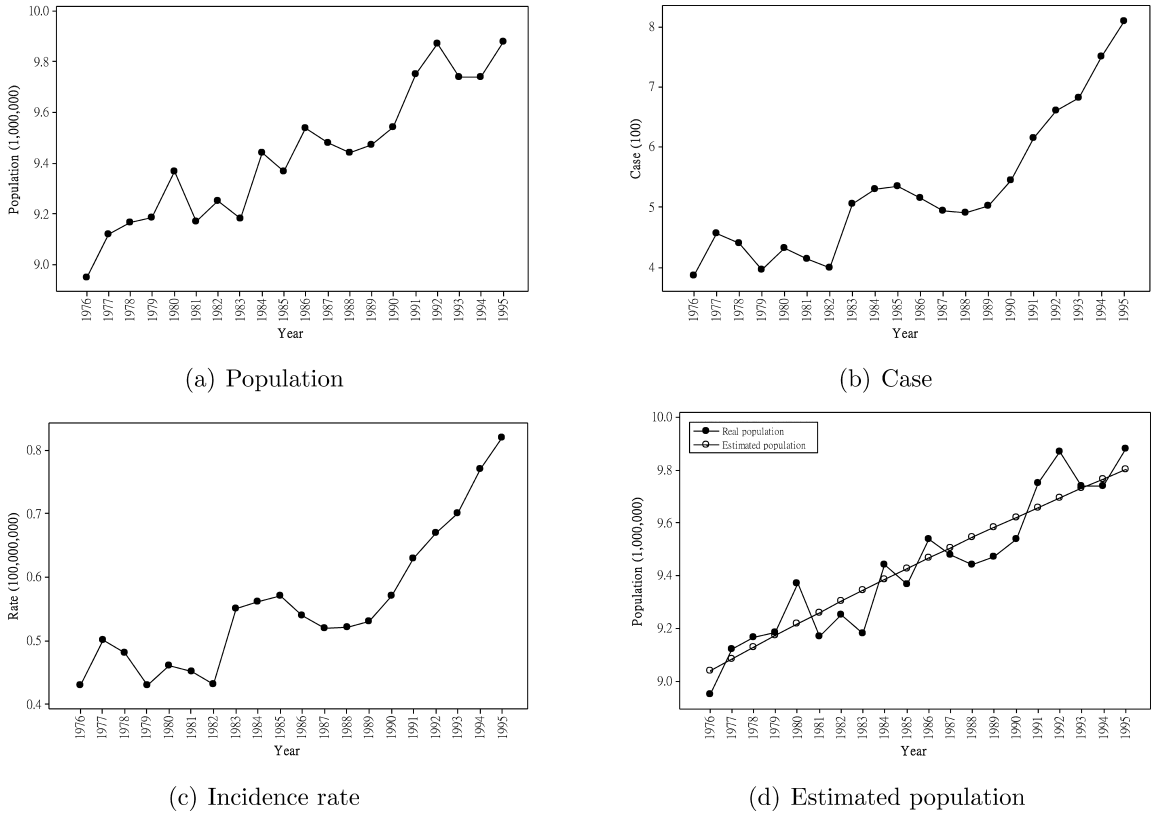
Before monitoring the incidence rate after 1983, we set $\alpha = 0.0027$ or equivalently $\mathrm{ARL}_0 = 370$. We fit a logistic model to the observed population sizes in Phase I using a nonlinear least square method (year 1975 is treated as time 0 and the population sizes are in the units of 1,000,000) and obtain $n_t = 5.856/\left[0.5 + \exp\{-(t + 86.295)/45.645\}\right]$ (Scenario I). Figure 3d shows the real population sizes and the expected population sizes estimated by the fitted logistic regression model. It indicates that in general, the natural character of the population growth can be well described by this logistic model. To show the adverse impacts of erroneously estimated population sizes, we further assume that the population sizes are constant with $n_t = 9.0$ (Scenario III) or uniformly distributed with $n_t \sim U(7.0, 10.0)$ (Scenario IV) for comparison.

Under the three different scenarios of population sizes discussed above, the control limits of the EWMAe chart are determined to be $\theta_0 + L\sigma_t$ with $L_{(\mathrm{I})} = 2.533$, $L_{(\mathrm{III})} = 2.547$, and $L_{(\mathrm{IV})} = 2.553$, respectively. Figure 4 plots the charting statistics (the solid curves connecting the dots) and the corresponding control limits (the dashed curves) of the EWMAG chart and the EWMAe charts under different population assumptions.

A significant increase in incidence rate can be observed from 1990. Therefore, an alarm should be issued as soon as possible after 1990. From the plots, it can be seen that the EWMAG chart exceeds its control limit in 1994 and remains above the control limit ever since. The EWMAe chart in Scenario I also triggers a signal in the same year. It makes sense that similar detection results are obtained from the two control charts as the natural population sizes are appropriately modeled in Scenario I. Conversely, the EWMAe chart issues a delayed OC signal in year 1995 in both Scenarios III and IV, which is caused by the inappropriate distributions of population sizes. The example demonstrates the usefulness of the EWMAG chart in reality because its detection capability does not depend on the correct estimation of population sizes, which is difficult to determine in advance.

6. CONCLUSIONS

Traditional control charts for monitoring Poisson count data with time-varying sample sizes have a severe shortcoming—they depend strongly on the knowledge of



(a) Population (b) Case

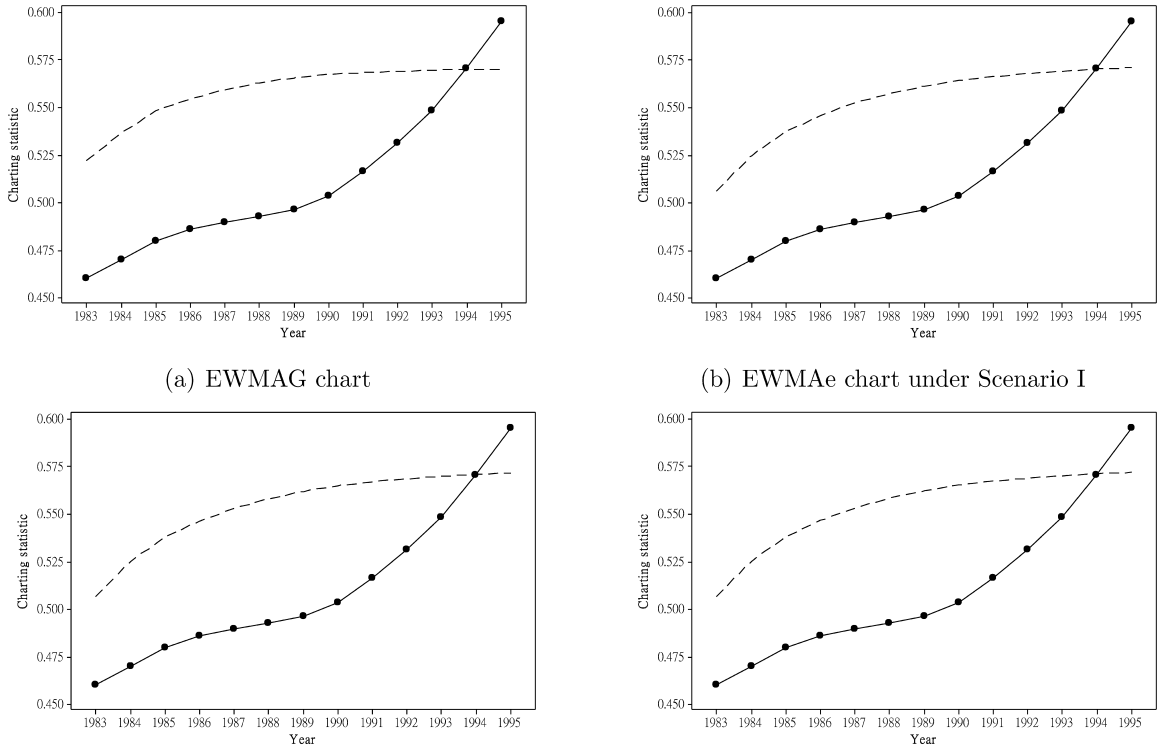
(c) Incidence rate (d) Estimated population

Figure 3. Female thyroid cancer incidence data and the expected population size estimated by the fitted logistic model. (a) Female population, (b) number of thyroid cancer cases, (c) incidence rate, and (d) estimated population sizes based on the fitted logistic model.

sample sizes before prospective monitoring starts. Such knowledge is seldom available in real applications. Our analysis shows that an inappropriate assumption or estimation of sample sizes may lead to poor run length performance of the traditional control charts. To overcome this shortcoming, we propose the use of probability control limits, which are determined based on sample sizes observed online, with the traditional control charts. The online time-varying control limit acts as a constraint on the conditional probability that the control chart statistic exceeds the control limit at present given that not a single alarm has been issued before. This ensures that a specified false alarm rate is achieved with identical signal probabilities at each step of Phase II. In this article, an EWMA-type control chart with the probability control limits, termed the EWMAG chart, is discussed in detail. The presented idea can be readily applied to any effective control charts such as the CUSUM chart. The proposed EWMAG chart is able to deliver robust and satisfactory IC and OC run length performance under various situations, as the simulation studies in this article have shown.

Future research may be extended in the following directions. First of all, recall that in this article we applied the proposed EWMAG chart to monitor the occurrence rate of adverse events assuming that the count of events follows a Poisson distribution when the sample size is observed online. Clearly, the EWMAG chart can be extended to more general cases in which the observations follow a conditional distribution given some related parameters/covariates whose information can be obtained and updated online as well. For example, when monitoring and predicting shopping quantity in retail data mining, the frequency of purchase and other demographics play an important role in determining the baseline purchasing frequency and quantity [16].

In addition, it is well recognized that the performance of the EWMA-type chart depends on the smoothing parameter λ, which is simply set to a constant value in this article. One strand of our ongoing research investigates how to sequentially determine optimal values of λ in the EWMAG chart. Moreover, the occurrence rates, which correspond to different sample sizes, are considered equally informative in the current study. That is, the EWMAG chart pays the same amount of attention to a ratio R_i based on a small at-risk n_i as it does to one based on a large n_i, even though the later is more informative in some cases. A control chart with sample-size-varying smoothing parameters would be more reasonable [19].



(a) EWMAG chart

(b) EWMAe chart under Scenario I

(c) EWMAe chart under Scenario III

(d) EWMAe chart under Scenario IV

Figure 4. The EWMAG chart and the EWMAe charts for monitoring the incidence rate of female thyroid cancer.

Finally, more research is required to extend our method to Phase I analysis, in which detection of outliers or change-points in a historical dataset and estimation of the baseline incidence rate is of great interest. It is also known that the performance of all control charts is affected by the amount of data in the reference dataset. Thus, the Phase I sample size required to ensure reasonable performance of the control charts with estimated parameters should be determined. Furthermore, future research needs to be directed toward developing a selfstarting version of the EWMAG chart which can simultaneously update parameter estimates and check for OC conditions [23].

ACKNOWLEDGMENTS

The authors would like to thank the Editor, Associate Editor, and two anonymous referees for their many helpful comments that have resulted in significant improvements to this article. Changliang Zou's research was supported by National Natural Science Foundation of China (Grants 11101306, 11001138, and 70931004), Foundation for the Author of National Excellent Doctoral Dissertation of PR China (Grant H0512101), and New Century Excellent Talents in University (Grant NCET-12-0276). Wei Jiang's research was supported by National Natural Science Foundation of China (Grant 71172131), Ministry of Education of China (Grant NCET11-0321), and Shanghai Pujiang Program. Fugee Tsung's research was supported by RGC Competitive Earmarked Research (Grants 620010 and 619612).

REFERENCES

[1] C.M. Borror, C.W. Champ, and S.E. Rigdon, Poisson EWMA control charts, J Qual Technol 30 (1998), 352–361.
[2] D. Brook and D.A. Evans, An approach to the probability distribution of CUSUM run length, Biometrika 59 (1972), 539–549.
[3] Y. Dong, A.S. Hedayat, and B.K. Sinha, Surveillance strategies for detecting changepoint in incidence rate based on exponentially weighted moving average methods, J Am Stat Assoc 103 (2008), 843–853.
[4] M. Frisén and J. De Maré, Optimal surveillance, Biometrika 78 (1991), 271–280.
[5] F.F. Gan, Monitoring Poisson observations using modified exponentially weighted moving average control charts, Commun Stat Simul C 19 (1990), 103–124.
[6] D.M. Hawkins and D.H. Olwell, Cumulative sum charts and charting for quality improvement, 1st ed., Springer-Verlag, New York, 1998.
[7] D.M. Hawkins, P. Qiu, and C.W. Kang, The changepoint model for statistical process control, J Qual Technol 35 (2003), 355–366.

[8] L. Huwang, Y.H. Wang, A.B. Yeh, and Z.S. Chen, On the exponentially weighted moving variance, Nav Res Logist 56 (2009), 659–668.
[9] T.L. Lai, Sequential changepoint detection in quality control and dynamical systems, J R Stat Soc B 57 (1995), 613–658.
[10] J.M. Lucas, Counted data CUSUMs, Technometrics 27 (1985), 129–144.
[11] T.M. Margavio, M.D. Conerly, W.H. Woodall, and L.G. Drake, Alarm rates for quality control charts, Stat Probab Lett 24 (1995), 219–224.
[12] Y. Mei, S.W. Han, and K-L. Tsui, Early detecting of a change in Poisson rate after accounting for population size effects, Stat Sin 21 (2011), 597–624.
[13] D.C. Montgomery, Introduction to statistical quality control, 2nd ed., Wiley, New York, 1990.
[14] K. Nancy, Hormone therapy and the rise and perhaps fall of US breast cancer incidence rates: Critical reflections, Int J Epidemiol 37 (2008), 627–637.
[15] J. Poloniecki, O. Valencia, and P. Littlejohns, Cumulative risk adjusted mortality chart for detecting changes in death rate: observational study of heart surgery, BMJ 316 (1998), 1697–1700.
[16] P.E. Rossi, R. McCulloch, and C. Allenby, The value of purchase history data in target marketing, Mark Sci 15 (1996), 321–340.
[17] A.G. Ryan and W.H. Woodall, Control charts for Poisson count data with varying sample sizes, J Qual Technol 42 (2010), 260–274.
[18] L. Shu, W. Jiang, and K.-L. Tsui, A comparison of weighted CUSUM procedures that account for monotone changes in population size, Stat Med 30 (2011), 725–741.
[19] L. Shu, W. Jiang, and K.-L. Tsui, A comparison of exponentially weighted moving average based methods for monitoring increases in incidence rate with varying population size, IIE Trans (in press).
[20] C. Sonesson and D. Bock, A review and discussion of prospective statistical surveillance in public health, J R Stat Soc A 166 (2003), 5–21.
[21] C.H. White and J.B. Keats, ARLs and higher-order run-length moments for the Poisson CUSUM, J Qual Technol 28 (1996), 363–369.
[22] W.H. Woodall, The use of control charts in health-care and public-health surveillance, J Qual Technol 38 (2006), 89−134.
[23] P.F. Zantek and S.T. Nestler, Performance and properties of Q-statistic monitoring schemes, Nav Res Logist 56 (2009), 279−292.
[24] C. Zou and F. Tsung, Likelihood ratio based distribution-free EWMA schemes, J Qual Technol 42 (2010), 174–196.
[25] Q. Zhou, C. Zou, Z. Wang, and W. Jiang, Likelihood-based EWMA charts for monitoring Poisson count data with time-varying sample sizes, J Am Stat Assoc 107 (2012), 1049–1062.