Research Article

Quality and Reliability Engineering International

(wileyonlinelibrary.com) DOI: 10.1002/qre.1248

Published online 16 August 2011 in Wiley Online Library

Statistical Monitoring of Time-to-Failure Data Using Rank Tests

Zhiguo Li,[a*†] Shiyu Zhou,[b] Crispian Sievenpiper,[c] and Suresh Choubey,[c]

A common type of reliability data is the right censored time-to-failure data. In this article, we developed a control chart to monitor the time-to-failure data in the presence of right censoring using weighted rank tests. On the basis of the asymptotic properties of the rank statistics, we derived the generic formulae for the operating characteristic functions of the control chart to show the relationship between type I error probability, type II error probability, sample size, and hazard rate change. We presented case studies to illustrate the design procedure and the effectiveness of the proposed control chart system. We also investigated and compared the performance of the proposed monitoring procedure with some available monitoring techniques for nonconformities. Copyright © 2011 John Wiley & Sons, Ltd.

Keywords: control chart; hazard rate; rank test; right censoring

1. Introduction

Because of the fast development of information technology, the critical failure events and maintenance events are often recorded and readily available in most modern engineering systems during their operations. For example, the system event log file of a printer or a computerized tomography (CT) machine records the time points at which a critical component of interest fails or a maintenance activity occurs. Figure 1(a) illustrates such a typical event log for a specific component starting at T_s and ending at T_e. In this figure, two types of events are recorded: (i) the failure and replacement of the component (represented as K) and (ii) the replacement of the component due to causes other than failure such as scheduled/preventive maintenance (represented as C). The event type C is also called the *censoring event* because it "disrupts" the measure of the operational life of the components. From these event logs, the following very useful information can be extracted:

1. Time between events is the time interval between two adjacent failure events as illustrated in Figure 1(b). Without loss of generality, the starting interval $[T_s, t_1]$ is also defined as a time between events.
2. A censored time interval is the time interval between a C event and its immediate preceding failure event or another C event. In other words, a censored time interval starts from either a K or a C event and ends at a C event.
3. An uncensored (or observed) time interval is the time interval between a failure K event and its immediate preceding failure event or C event. In other words, an uncensored time interval starts from either a K or a C event and ends at a K event.

The extracted information of the censored and uncensored time intervals (called *time-to-failure data*) can be used to investigate the occurrence risk of the failure event of the component. In practice, if there is a significant increase of the occurrence risk of failure event over time, we know the performance of the machine degrades. This degradation could be due to multiple root causes, such as the changes in the usage pattern, the changes in the environment, the degradation of the machine as a whole, and so forth[1]. Obviously, it is very beneficial to detect the changes in the occurrence risk of failure and to provide the guidelines for reliability improvement.

This article develops a control chart technique to detect the change in the occurrence risk of failure event in the presence of right censoring. In the literature, some techniques have been developed to monitor the key events, such as the failure of machines/components or the mortality of patients, of an industrial process or a healthcare procedure. These techniques can be roughly classified into two categories: (i) the risk-adjusted (RA)[2,3] monitoring procedures and (ii) the non-risk-adjusted (NRA) monitoring procedures. To reduce the false alarms, the methods in the RA monitoring procedures take into account the fact that the subjects are

[a]IBM T. J. Watson Research Center, 1101 Kitchwan Rd. Route 134, PO Box 218, Yorktown Heights, NY 10598, USA
[b]Department of Industrial and Systems Engineering, The University of Wisconsin–Madison, Madison, WI 53706, USA
[c]Global Service Technology, GE Healthcare, Waukesha, WI53186, USA
*Correspondence to: Zhiguo Li, Department of Industrial and Systems Engineering, The University of Wisconsin–Madison, Madison, WI 53706, USA.
†E-mail: Zhiguo.li.wisc@gmail.com

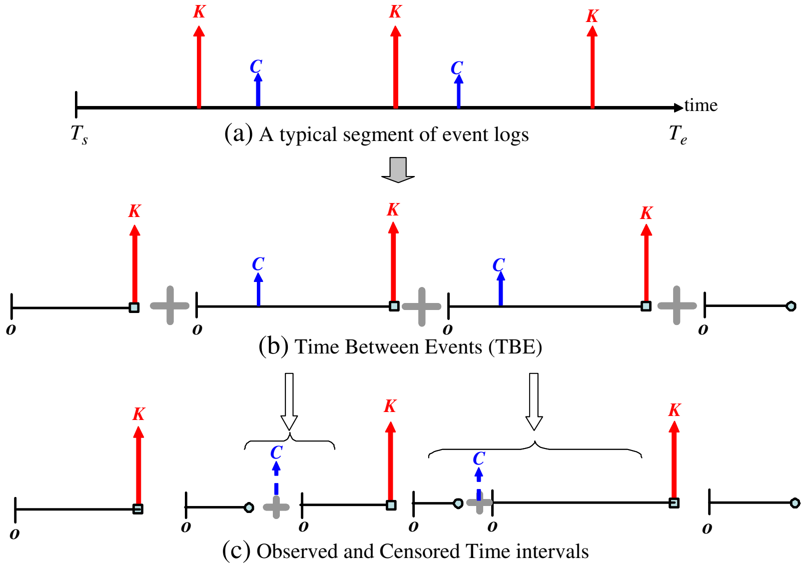

Figure 1. Decomposition of an event sequence time interval

heterogeneous, and thus the risk of failure (or mortality) for each subject depends on both the degradation process (or healthcare procedure) and the risk factors associated with the unique characteristics of each subject. In the monitoring procedures, a regression model (called *the risk adjustment model*), such as the accelerated failure time regression model, is fitted to the in-control data to link the occurrence risk of key event to the subject characteristics. Some proposed examples include the RA Bernoulli cumulative sum (CUSUM) control chart[4] and the RA survival time CUSUM chart[5], which are used to monitor the binary outcomes (e.g. whether a patient can survive after a specified time) and the survival times.

When the subjects under investigation can be viewed as homogenous, for example, in Figure 1, failures occurred to the component for the same machine, non-risk-adjusted control charts are used to monitor the occurrence risk of event of interest. In the statistical process control of manufacturing production, some techniques have been proposed to monitor the occurrence rate of nonconforming items, which can be viewed as the failure event K in Figure 1. These methods can be roughly classified into count-based methods and interarrival time-based methods. With the Poisson model for the nonconforming items, Poisson counts are calculated by partitioning the time interval into regular subintervals, and then the number of events in each subinterval is counted. The Shewhart control charts are proposed to monitor the count data, such as c and u charts. Lucas[6] designed the Poisson CUSUM chart, and Borror *et al.*[7] developed the Poisson exponentially weighted moving average (EWMA) chart for Poisson data. The interarrival time-based methods provide an alternative way to monitor the occurrence rate of the nonconforming items. The exponential CUSUM chart was firstly proposed by Vardeman and Ray[8] to monitor the Poisson rate, and Gan[9] developed a simple procedure to design an optimal exponential CUSUM chart. Gan[10] also studied the EWMA charts for exponential data. Moreover, the Shewhart control charts can also be applied to exponential interarrival times. As an example, Nelson[11] recommended using a power transformation of the exponential data, and then the Shewhart charts (hereafter called the *transformed time-between-events chart*) can be used for individual measurements to monitor the process. In general, these NRA control charts can be used to monitor the occurrence rate of nonconformities/defects, but they cannot be applied directly to the right censored data.

If a nontrivial portion of the data is right censored, then it is of importance to take censoring into consideration when monitoring the occurrence of failure event. Steiner and MacKay proposed the Shewhart type $\bar{X}$ and S charts for monitoring processes with observations censored at a fixed level[12] and an EWMA control chart in the case of right censoring due to competing risks[13]. Both methods use the conditional expected value (CEV) weight with the distributional assumption of the data. Furthermore, in the EWMA CEV chart, the same distribution (e.g. the normal distribution) is assumed for both the failure and the censoring events (called *competing risks* in the article) to derive the CEV weights.

In summary, all the existing NRA control chart techniques were developed on the basis of specific distributional assumptions of the data. However, in many cases, the time-to-failure data, including both failure and censoring events, does not follow a specified distribution. It is desirable to have a monitoring procedure for the occurrence rate of the failure event without distributional assumptions.

In this article, we propose a new non-risk-adjusted monitoring procedure to monitor the occurrence rate of right censored failures without assuming specific distributions for the time-to-failure data. In this procedure, we will monitor the weighted rank statistic, which is a nonparametric statistic for comparing hazard rates of two populations of failure times. The hazard rate can be viewed as the instantaneous failure rate at any time point. Readers are referred to the article by Harrington and Fleming[14] for details of the rank tests for censored survival data, and a description of the rank statistic can also be found in the following section. The major advantage of the proposed method is that we do not need to assume any specific distribution for the data. Furthermore, the data censoring is handled naturally by a nonparametric estimator of the cumulative hazard function that is used in the test statistic. However, the use of the rank test statistic does not require censored data. Thus, in the case without censoring, it can still outperform the conventional NRA charts when the data does not follow a specific distribution, such as the Poisson, the exponential, or the geometric distribution[15], required by those charts.

The remainder of this article is organized as follows. We first motivate the procedure by introducing the families of weighted rank tests for the equality test of hazard rates of two populations of failure times. After that, we present the monitoring procedure and illustrate its application. The generic formulae for the operating characteristic (OC) function of the control chart are derived to show the relationship between type I error probability, type II error probability, sample size, and hazard rate change on the basis of the asymptotic properties of the rank statistics. Finally, numeric case studies are implemented to show the effectiveness and investigate the performance of the proposed monitoring procedure.

2. Description of the monitoring statistic

As stated in the Introduction, we will consider monitoring time-to-failure data, which is also called *survival data* in survival analysis[16]. Hereafter, we assume that the censoring times (i.e. occurrence times of event "*C*") are independent of the failures. This assumption is commonly used in practice for replaceable components/subsystems[17].

Let T be the random variable representing the uncensored time to failure. If the density function of T, $f(t)$, exists, it is known that the survival function of T is

$$S(t) = \Pr(T > t) = \int_t^{+\infty} f(x)dx \tag{1}$$

The hazard function (or simply called *hazard rate*), also called *conditional failure rate function* in reliability, is defined as

$$h(t) = \lim_{\Delta t \to 0} \frac{\Pr(t \le T < t + \Delta t | T \ge t)}{\Delta t} = \frac{f(t)}{S(t)} \tag{2}$$

We can interpret the hazard function as the "instantaneous" probability that the failure event occurs at time t, given that no failure event occurs before t. Thus, $h(t)\Delta t$ can be viewed as the "approximate" probability a failure event will occur in a small time interval between t and $t+\Delta t$. If $h(t)$ is known, we can calculate the function $S(t)$ according to the following equation,

$$S(t) = \exp\left[-\int_0^t h(x)dx\right] = \exp[-H(t)] \tag{3}$$

where $H(t) = \int_0^t h(x)dx$ is the cumulative hazard function.

Now we can use the hazard rate $h(t)$ to measure the occurrence risk of the failure event. Then our task is to test if the hazard function changes or not during certain period. Specifically, consider two data sets, called the *historical data set* and the *monitoring data set*. Further, let $h_1(t)$ and $h_2(t)$ represent the hazard functions of these two data sets, respectively. To test the changes of the hazard function over the two data sets, we can set up the following hypotheses (called the *two-sample test*):

$$H_0: h_1(t) = h_2(t) \text{ for all } t \le \tau, \text{ versus}$$

$$H_1: h_1(t) \ne h_2(t) \text{ for some } t \le \tau,$$

where τ can be selected as the smaller value of the largest failure time in each data set. Such a selection of τ is because the nonparametric estimates of survival function $S(t)$ and cumulative hazard function $H(t)$ are not defined after τ, and thus we cannot compare the two hazard functions when $t > \tau$ [16]. It needs to be pointed out that to use this test, we need two data sets (the historical and the monitoring data sets), and thus the exact value of τ will be known after both data sets are collected. As the event sequence grows over time, multiple monitoring data sets will be collected, and each monitoring data set will be plotted as a point on the control chart. The statistic used for this hypothesis test is presented as follows.

Let $t_1 < t_2 < \ldots < t_D$ be the distinct uncensored failure times (i.e. time points when failure occurs) in the pooled sample from both data sets, that is, the failure events occur at D distinct times. Hereafter, the terms *uncensored failure time* and *failure time* are used interchangeably. At time point t_i, we observe d_{ij} failure events in the jth data set out of Y_{ij} failure times that are no smaller than t_i, where i=1, 2,. . ., D is the index of occurrence time, and j=1, 2 is the *index of* data set. Let $d_\bullet(t_i) = d_{i\bullet} = \sum_{j=1}^{2} d_{ij}$ and $Y_\bullet(t_i) = Y_{i\bullet} = \sum_{j=1}^{2} Y_{ij}$ denote the number of failure events at time t_i, i=1, 2, . . ., D and the number of failure times not smaller than t_i in the pooled sample. A simple example is illustrated in Figure 2 to explain the notations stated earlier.

Assume we have four uncensored and two censored failure times in the pooled sample from the historical and monitoring data sets and there are four distinct times, $t_1 < t_2 < t_3 < t_4$, at which failure events occur (t_0 from the monitoring data set was not included in the set of distinct times t_1 through t_4 because it is the length of a censored time interval). Note that one uncensored and one

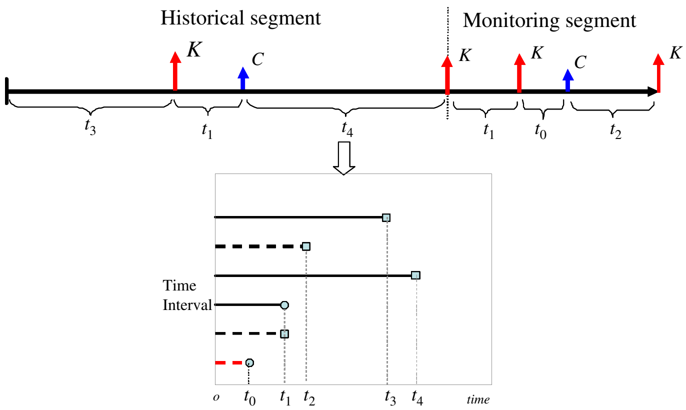

Figure 2. An illustration of notations

censored failure times (t_1) are the same. In the figure, the solid lines are from group 1, whereas the dashed lines are from group 2. The symbol "□" represents an uncensored failure time, and the symbol "o" means a censored failure time. At time t_1, $d_{1.}$=1 because one failure event occurs and $Y_{1\bullet}$=5 because we have five uncensored and censored failure times not smaller than t_1. Similarly, we know d_{11}=0, Y_{11}=3, d_{12}=1, and Y_{12}=2. At t_3, the values of $d_{3\bullet}$ and $Y_{3.}$ become 1 and 2. Moreover, we have d_{31}=1, Y_{31}=2, d_{32}=0, and Y_{32}=0.

The nonparametric two-sample test statistic, on the basis of the Nelson–Aalen estimator of the cumulative hazard function[16], is as follows:

$$Z_j(\tau) = \frac{\sum_{i=1}^{D} W(t_i)\left\{d_{ij} - Y_{ij}\frac{d_{i\bullet}}{Y_{i\bullet}}\right\}}{\sqrt{\sum_{i=1}^{D} W(t_i)^2 \frac{Y_{ij}}{Y_{i\bullet}}\left(1-\frac{Y_{ij}}{Y_{i\bullet}}\right)\left\{\frac{Y_{i\bullet}-d_{i\bullet}}{Y_{i\bullet}-1}\right\}d_{i\bullet}}}, j = 1, 2 \tag{4}$$

where $W(t_i)$ is a nonnegative weight function and $W(t_i)$ = 0 whenever $Y_{i.}$ is zero. The statistic has a standard normal distribution for large samples under H_0 with $Z_1(\tau)=-Z_2(\tau)$. Thus, the null hypothesis will be rejected when $|Z_2|\geq Z_{\alpha/2}$ (or $|Z_1|\geq Z_{\alpha/2}$) at a significance level of α. In this article, we will adopt this statistic as the statistic to monitor. Intuitively, this statistic tries to compare the ratios d_{ij}/Y_{ij} and $d_{i\bullet}/Y_{i\bullet}$, the hazard rates for the *j*th data set and the pooled data set consisting of both data sets (*j*=1, 2), at each distinct time t_i, *i*=1,2,.. *D*, and sum the differences together.

There are multiple choices of the weight function $W(t_i)$. A trivial weight function, $W(t_i)$ = 1 for all t_i, leads to the so-called *log-rank test*. However, other weight functions, such as (*i*) $W(t_i)=g(Y_{i.})$, where $g(\cdot)$ is a continuous and square integrable function on [0, 1] and (*ii*) $W(t_i) = \tilde{S}(t_i)^{\rho}$ where $0\leq\rho\leq1$ and $\tilde{S}(t_i) = \prod_{t_k\leq t_i}\left(1-\frac{d_{k\bullet}}{Y_{k\bullet}+1}\right)$, can also be used. These weight functions lead to the Tarone–Ware family[18] of tests and the Harrington–Fleming family of tests[14]. Two-sample tests with different weight functions in Equation (4) have the optimal power for different alternative hypotheses. As an example, the log-rank test is locally optimal for proportional hazards alternatives, and it is widely used, for instance, to compare survival curves in the design of clinical experiments and analysis of clinical data, where the hazard rate of the group of patients receiving new treatment is assumed to be proportional to that of the control group.

When designing the monitoring procedure with the two-sample test statistic, we need to specify the sample size (i.e. the number of *uncensored* failure times) n_2 of the subgroups of monitoring data and the sample size n_1 in the historical data set that is used as the baseline in the two-sample test. Each subgroup (monitoring data set) is compared with the historical group, and the corresponding value of the rank statistic is plotted as a point on the charts. In the chart, each subgroup contains the same number of failure times. This is different from most of the count-based charts such as *c*-chart, where we fix the length of the sampling time interval, but count the number of events in each interval.

In the next section, we shall describe the design procedure for the rank test chart and derive its OC function.

3. Design of the rank test control chart

To establish a control chart based on the rank test, we first need to select the proper weight function. Because different weight functions will provide optimal tests for different alternative hazard function, it is desirable to know what kind of alternative hazard function we need to monitor and then we can select the appropriate test type. For example, the log-rank test is optimal in the case of proportional hazards, although it can be used in nonproportional case as well. After the weight function is selected, we need to decide the following parameters for the control chart: (i) the type I error probability (α); (ii) the type II error probability (β); (iii) the upper control limit (UCL)—once the value of α is specified, the value of UCL of *upper* rank test chart is UCL=Z_α, whereas UCL=$Z_{\alpha/2}$ and LCL=$-Z_{\alpha/2}$ for the two-sided chart; (iv) the baseline hazard rate function $h_1(t)$, which can be estimated using the historical data; (v) the intended alternative hazard rate function $h_2(t)$, which can be decided according to the practical requirements, for example, if one wants to monitor the proportional hazard, then one can specify the hazard ratio k and thus $h_2(t) = kh_1(t)$; and (vi) the sample sizes n_1 and n_2, where n_j is the number of *uncensored* failure times in the jth group, j=1, 2 for historical group and monitoring group.

To calculate the sample sizes, we need to derive generic formulae on the basis of the asymptotic properties of the statistics in Equation (4) for the OC functions of the monitoring procedure to link the type II error probability (β) with the type I error probability (α), the sample size (n_1 and n_2), and the hazard rate change. The out-of-control average run length (ARL) can be calculated on the basis of the formulae to help the practitioners design the control charts.

3.1. OC function

In this section, we will study the OC of the monitoring procedure developed on the basis of the two families of rank statistics: the Tarone–Ware family and Harrington–Fleming family. To derive the OC function of the proposed procedure, we will investigate the asymptotic properties of the statistic under the null hypothesis and the alternative. For this goal, we first introduce some notations:

1. $S_j(t)$ and $G_j(t)$ are the survival functions for the uncensored and censored failure times, respectively, for the jth data set, j=1, 2. In addition, denote $G_j(s-)$as $\Pr(C_j \geq s)$, where C_j is the random variable representing the censored time in the jth group.
2. n_j is the number of uncensored failure times in the jth group and let n be n_1+n_2, p_j be n_j/n, and thus $\sum_{j=1}^{2} p_j = 1$.
3. Denote $y_j(s)$as $p_jS_j(s)G_j(s-)$, j=1, 2 and $y_\bullet(s) = \sum_{j=1}^{2} y_j(s)$.
4. Let $w(s)$ be $g(y_\bullet(s))$ for the Tarone–Ware family and $w(s)$ be $S(s)I(y_\bullet(s) > 0)$ for the Harrington–Fleming family. There are two special cases of the function $w(s)$: $w(s)=I(y_\bullet(s) > 0)$ for the log-rank test, whereas $w(s)=y.(s)$ for the Gehan–Breslow test. Here $I(\bullet)$ is the indicator function.

Using the previously mentioned notations and assuming the failure times are independent, we have the following result regarding the OC function and the proof of this result is presented in the Appendix. This result can be used to calculate the sample size needed when designing the rank test control chart.

Under the null hypothesis and alternative,

$$H_{0:} h_1(t)=h_2(t) \text{ for all } t \leq \tau$$

$$H_{1:} h_1(t) \neq h_2(t) \text{ for some } t \leq \tau$$

For the Tarone–Ware family and Harrington–Fleming family of tests, we have

$$n = \left[\left(Z_{\alpha/2} \cdot \sigma_0 + Z_\beta \cdot \sigma_1\right)/\zeta\right]^2, \tag{5}$$

with $\sigma_0^2 = \int_0^\infty w^2(s) \frac{y_1(s)y_2(s)}{y_\bullet(s)} h_1(s)ds$,

$$\sigma_1^2 = \int_0^\infty w^2(s) \frac{y_1(s)y_2(s)}{y_\bullet^2(s)} [y_1(s)h_2(t) + y_2(s)h_1(t)]ds \text{, and}$$

$$\zeta = \int_0^\infty w(s) \frac{y_1(s)y_2(s)}{y_\bullet(s)} [h_2(s) - h_1(s)]ds, \tag{6}$$

where α and β are type I and type II error probabilities of the test, and Z_α and Z_β represent the upper percentage points of standard normal distribution.

Equation (5) is the expression of the relationship between sample size n, type I and type II error probabilities, and hazard rates $h_1(t)$ and $h_2(t)$. Intuitively, this equation is similar to the sample size formula for comparing the means of two normally distributed samples with equal standard deviation, $n=[(Z_{\alpha/2}+Z_\beta) \cdot \sigma/\Delta]^2$, where σ is the standard deviation and Δ is the expected mean difference. In Equation (5), σ_0 and σ_1 can be viewed as the standard deviation of the asymptotic normal distributions under H_0 and H_1, and ζ can be viewed as the distance of the mean between these two distributions. Several remarks of Equation (5) are listed as follows:

- Equation (5) is an approximate sample size formula. In practice, numerical integration is used for calculating σ_0^2, σ_1^2, and ζ in Equation (5). Because the survival function $S(t)$ (and $y(t)$) often approaches 0 quickly, a large upper limit could be used to replace ∞ in calculation of integration without resulting in a significant bias.
- When the difference in the hazard functions under H_1 is small, that is, $h_1(s) \approx h_2(s)$, people often assume that σ_1^2 is approximately the same as σ_0^2, and thus we have

$$n = \left[\left(Z_{\alpha/2} + Z_\beta\right)\sigma/\zeta\right]^2 \tag{7}$$

- The values of p_1 and p_2 can be selected arbitrarily from (0, 1) in theory. However, we usually select a larger value for p_1 ($p_1 > 0.5 > p_2$), considering the fact that the sample size of the baseline (in-control) group usually is larger because a large amount of historical data are often available. Hereafter, we will only consider the case of $p_1 > 0.5 > p_2$ in this article.
- For an α level test of the one-sided alternative H_A: $h_1(t) < h_2(t)$, for some $t \leq \tau$, the sample size formulae (Equation (5)) can be modified by simply replacing $Z_{\alpha/2}$ with Z_α and thus $n_A=[(Z_\alpha \cdot \sigma_0+Z_\beta \cdot \sigma_1)/\zeta]^2$. With this result, we can set up the *one-sided upper rank test chart*. The same result holds for the one-sided alternative H_A: $h_1(t) > h_2(t)$ for some $t \leq \tau$. In this study, we will focus on the upper

rank test chart because people are more interested in detecting the increased hazard rate in practice. In the following sections, all numerical case studies are under the alternative

$$H_{A:}\, h_1(t) < h_2(t) \text{ for some } t \le \tau.$$

Please note that even if a nonparametric statistic in Equation (4) is used for comparing the hazard functions, the distributions of the failure times and censoring times should be specified when the properties of the statistic under the alternative need to be investigated, as shown in Equation (5). As we state previously, the log-rank test is locally most powerful for proportional hazards alternatives, and it is widely used to compare survival curves. Therefore, we would pay special attention to the log-rank test. We have the following result.

Under the null and alternative hypotheses in Equation (5), for the two-sample *log-rank test* with the proportional hazards $h_2(t)=k{\cdot}h_2(t)$ for $t \le \tau$, where $k \in (0,\infty)$ is a constant, we have

$$n = \left[\left(Z_{\alpha/2}{\cdot}\sigma_0^* + Z_\beta{\cdot}\sigma_1^*\right)/\zeta^*\right]^2 \tag{8}$$

with

$$(\sigma_0^*)2 = \int_0^\infty I(y_\bullet(s) > 0)\frac{y_1(s)y_2(s)}{y_\bullet(s)}h_1(s)ds, \left(\sigma_1^*\right)^2 = \int_0^\infty I(y_\bullet(s) > 0)\frac{y_1(s)y_2(s)}{y_\bullet^2(s)}[ky_1(s) + y_2(s)]h_1(s)ds'$$

and

$$\zeta^* = (k-1){\cdot}\int_0^\infty I(y_\bullet(s) > 0)\frac{y_1(s)y_2(s)}{y_\bullet(s)}h_1(s)ds'$$

where $w(s)=I(y.(s)>0)$ and $I(\cdot)$ is the indicator function.

This result can be easily derived from Equation (5), and the proof is omitted here. From Equation (8), we can see that if the difference in hazard functions under H_1 is small, then

$$n = \left[\frac{\left(Z_{\alpha/2} + Z_\beta\right)}{(k-1){\cdot}\sigma_0^*}\right]^2 \tag{9}$$

Using Equations (5) and (8), we can calculate the sample size n with given values of α and β after the hazard functions $h_1(t)$ and $h_2(t)$ in two groups (thus $S_1(t)$ and $S_2(t)$ are known) and the proportions p_1 and p_2 are specified. Also, we can use these equations to calculate the power $(1-\beta)$ of our monitoring procedure when the other parameters are specified.

As an example, Figure 3 plots the OC curves for the two-sample log-rank test with α=0.01 and p_1=0.8 and p_2=0.2. The failure times in the historical data is assumed to follow $W(4\times10^{-4}, 2)$, where $W(a,b)$ represents the Weibull distribution with the scale parameter a and shape parameter b and its hazard rate is $h_0(t)=abx^{b-1}$. The Weibull distribution[19] is widely used in the reliability analysis because of its flexibility in modeling the failure behavior. Furthermore, we assume the failure times in monitoring data follow $W(k{\cdot}4\times10^{-4}, 2)$, where the range of hazard ratio k is from 1 to 5. Please note that, in this study, the assumed alternatives are only used as examples to show how to determine an acceptable sample size for given type I and type II error possibilities, which does not indicate that the proposed rank-test chart is designed for a specific alternative. It is assumed that no censoring exists in the data. The solid and dotted lines represent the methods of Equations (8) and (9). From the top to the bottom, the sample sizes are 20, 40, 60, 80, and 100.

From the plot we can see, when the hazard ratio k is close to 1 and sample size is large, the results of Equations (8) and (9) are quite close. However, when the hazard ratio is large, which means the alternative is quite different from the null hypothesis, the discrepancy between these two methods is large. The OC curves are derived on the basis of the asymptotic properties of the statistic; therefore, the accuracy of these formulae should be checked. This problem is investigated in the numerical studies.

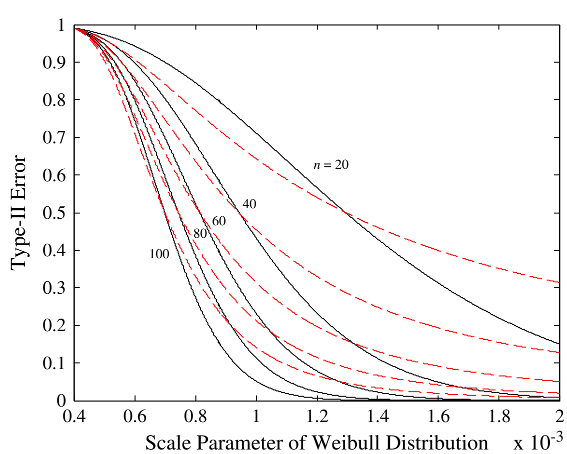

Figure 3. Operating characteristic curves for the two-sample log-rank test with α =0.01

3.2. A design example for the censored failure times

A typical design procedure is illustrated as follows. To design an upper rank test chart to detect the increased proportional hazard ratio k=2 with ARL_0=100, we use the *log-rank test* to monitor the hazard rate in this case. In this study, the data in control are assumed to follow $W(4\times10^{-4}, 2)$, whereas the distribution for out-of-control data is $W(8\times10^{-4}, 2)$. The hazard function $h(t)=abx^{b-1}$ can be obtained accordingly. The censoring distribution is Exp(0.005) for both in-control and out-of-control data, and there is approximately 20% censoring in the in-control data. In practice, the in-control distribution and the censoring distribution can be estimated on the basis of the historical data under normal working conditions, whereas the practitioners will specify the out-of-control parameters to monitor and thus the sample sizes can be calculated. The design steps are as follows:

- The desired in-control ARL_0 is 100 (α=0.01).
- The intended alternative hazard ratio k *is* 2.
- Given β=0.25, Equation (9) gives the sample sizes n_1 and n_2 as 100 and 25. They are used in the following analysis.
- Plot the values of the two-sample statistic in Equation (4) against the subgroup index in the control chart with UCL=z_α. Finally, the rank test chart is illustrated in Figure 4(a). There are a total 92 subgroups, and each subgroup consists of 25 failure times.

The one-sided *c*-chart and the transformed time-to-event chart with the same type I error possibility[20] are also plotted in Figure 4. For the *c*-chart and the transformed time-to-events charts, both observed and censored times are linked together to form a series of events (i.e. nonconformities) over time because they cannot handle censoring. That is, the censoring events are viewed as failures in these two charts. In addition, the sampling time interval is taken as $n_2\cdot\bar{y}_1$, where $\bar{y}_1$ is the average of times in the simulated out-of-control data so that the count of events in a sampling interval is comparable with the sample size of the rank test chart. The transformed time-to-event charts consist of the moving range chart (Figure 4(c)) and the chart for individuals (Figure 4(d)). Notice that we use the index of failure time in Figures 4(c) and 4(d), and for the sake of clarity, only a portion of all failure times is plotted in the figure (the change occurred after the 60th point). If all points are plotted on the control chart, the figures will become hard to read.

It can be seen that the performance of the rank test chart is the best, whereas the transformed time-to-event charts cannot detect the change. The study is repeated 10,000 times, and the ARL_1 values of the rank test chart and the *c*-chart are 1.345 and 8.673, respectively. To make a fair comparison, the out-of-control average time to signal (ATS; i.e. the expected time required for a chart to signal an alarm from the start of the shift) values are also calculated. The ATS values for the rank test chart and the *c*-chart are 962.99 and 6190.85, respectively.

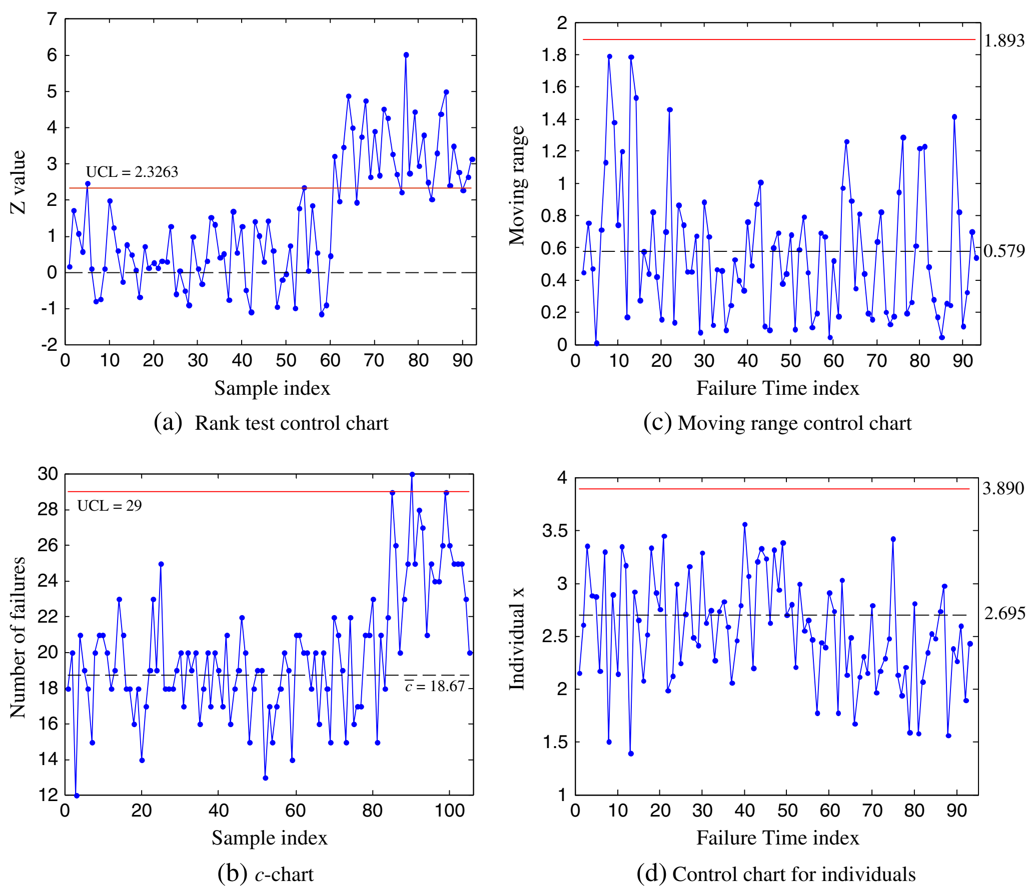

Figure 4. Control charts for event sequence data

4. Numerical studies

To show the performance of the proposed control charting method as well as the accuracy of derived sample size formulae, we carry out the following numerical studies.

4.1. *Performance of rank test control chart*

In this section, we implement numerical study to investigate the performance of rank test chart for the Weibull and exponential data in terms of out-of-control ARL_1 and ATS and compare it with the *c*-chart, exponential CUSUM chart, and exponential EWMA chart. Two scenarios are considered here: (i) no censoring exists in the data and (ii) some failure times are censored. The corresponding results are illustrated in the following sections.

Gan[10] designed the procedure of one-sided and two-sided exponential EWMA charts. In his article, the author compared ARL_1 values of lower-sided CUSUM and EWMA control charts with respect to out-of-control exponential means. For exponential distribution Exp(λ) with mean μ=1/λ, the lower-sided exponential EWMA chart optimized for μ=0.20 in the steady state have the following parameters: $\lambda_q = 0.33$, $h_q = 0.25$, and $q_0 = 0.98$, whereas the lower-sided exponential CUSUM chart under the same conditions has $k_T = 0.40$, $h_T = 1.24$, and $T_0 = -0.14$ (for details about these parameters, refer to Table 4 in Gan[10]). In this work, the in-control ARL value is set to 500, and thus α should be 1/500=0.002 for the rank test chart. In the following sections, we would compare the performance of the rank test chart with the performance of the *c*-chart, the exponential CUSUM chart, and the EWMA chart in two cases of (i) the Weibull data and (ii) the exponential data in terms of two metrics, ARL and ATS, which are averaged values from multiple runs of simulation. In each run, a historical data set and the multiple monitoring data sets are used to determine the run length or time to signal for each control chart, and finally these run lengths or time-to-signal values are averaged to estimate the values of ARL and ATS.

4.1.1. Data without censoring. In the first scenario, we assume there is no censoring in the data. As stated previously, the sample size for the rank test chart calculated through Equation (8) or Equation (9) for the Weibull distributions $W(a, b)$ and $W(ka, b)$ are the same as the sample size for Exp(λ) and Exp($k\lambda$), regardless of the specific values of a, b, and λ. Therefore, without loss of generality, we can calculate the sample sizes using Exp(1) and Exp(k).

Given β=0.20, the sample sizes of monitoring group n_2 for p_1=0.7, 0.8, and 0.9 calculated through Equations (8) and (9) derived for the log-rank test are listed in Table I, where methods I and II represent Equations (8) and (9), respectively. These results are for both the Weibull and the exponential data when k=1.5, 2, 2.5, 3, and 4. From Table I, we can see that when k or p_1 increases, the sample size will decrease. Compared with method I, method II tends to give smaller sample sizes. This is because method II uses the variance of the rank statistics under H_0 to approximate the variance under H_1, and the latter is larger than the former according to Equation (8).

As an example, we use the sample sizes calculated when p_1=0.8 to construct the rank test chart. The Weibull and the exponential data are generated, and the historical group is assumed to follow $W(1, 2)$ or Exp(1). For all control charts, the study is repeated 10,000 times (including the simulation of the preliminary sample, if any) for each different value of k, and finally the average values of ARL_1 and ATS for each control chart are computed. In both cases, for the *c*-chart, we simulated a preliminary sample of $100 \times n_2$ failure times on the basis of which the mean of the distribution can be estimated, and again the sampling time interval is taken as $n_2 \cdot \bar{y}_1$. The total number of failure times in the in-control data under the distribution $W(1, 2)$ or Exp(1) is $n_1 + 1000 \times n_2$, and then we induce a change in the distribution of the failure times so that the new distribution is $W(k, 2)$ or Exp(k).

The final results for the Weibull and exponential data are shown in Tables IIA and IIB. In Tables IIA(a) and Table IIB(a), the number of iterations is only 1000 because the ARL_1 and ATS values are very large for the CUSUM and EWMA control charts. To make the results comparable, the third column of Table IIA lists the values of ARL_1 multiplied by the sample size n_2. We can use the left panel of Table IIA to compare the performance of the rank test chart and *c*-chart. The right panel is used to compare the rank test chart with the exponential CUSUM and EWMA control charts.

Although the performances of the rank test control chart for the Weibull and exponential data are quite similar, we can see there are some apparent differences for the other three charts because they are developed under the assumption of exponential distribution for interarrival times. For exponential data, the performance of the rank test chart is a little worse than the performance of the other three charts; however, it is much better in the case of Weibull data. Method II has larger ARL_1 values compared with method I because of smaller sample sizes, whereas the ATS values from these two methods are close.

Table I. Sample sizes n_2 calculated through methods I and II (α=0.002, β=0.20)

	p_1					
	0.7		0.8		0.9	
k	I	II	I	II	I	II
1.5	119	110	108	99	99	90
2	41	36	38	33	36	30
2.5	24	20	23	18	21	17
3	17	13	16	12	15	12
4	11	8	11	8	10	7

Table IIA. ARL_1 of four types of control charts

k	Rank test chart I	Rank test chart II	c-chart I	c-chart II	Rank test chart $\times n_2$ I	Rank test chart $\times n_2$ II	Exponential CUSUM	Exponential EWMA
(a) Weibull data								
1.50	1.28	1.38	12.01	14.44	137.96	136.43	37718.09	13188.59
2.00	1.26	1.43	10.76	13.81	47.95	47.27	10848.08	2390.80
2.50	1.23	1.57	9.10	13.61	28.22	28.18	2365.92	718.36
3.00	1.24	1.64	7.79	15.21	19.81	19.63	765.84	318.22
4.00	1.22	1.86	5.35	14.01	13.45	14.91	129.38	99.59
(b) Exponential data								
1.50	1.28	1.36	1.14	1.20	137.93	134.68	96.69	76.42
2.00	1.25	1.43	1.16	1.24	47.39	47.18	34.99	29.54
2.50	1.23	1.54	1.15	1.30	28.23	27.65	18.85	17.54
3.00	1.25	1.66	1.16	1.35	20.02	19.87	12.92	12.92
4.00	1.21	1.81	1.15	1.27	13.26	14.51	8.66	9.31

Table IIB. ATS of four types of control charts

k	Rank test chart I	Rank test chart II	c-chart I	c-chart II	Exponential CUSUM	Exponential EWMA
(a) Weibull data						
1.50	99.81	98.74	939.03	1034.58	27292.21	9542.04
2.00	30.08	29.59	256.65	285.97	6797.11	1498.18
2.50	15.80	15.79	117.74	137.70	1325.30	402.39
3.00	10.13	10.03	64.15	93.78	391.72	162.93
4.00	5.98	6.61	26.39	50.00	57.31	44.09
(b) Exponential data						
1.50	91.97	89.84	83.01	79.86	64.46	50.97
2.00	23.66	23.59	22.51	21.00	17.53	14.80
2.50	11.29	11.04	10.95	9.76	7.55	7.03
3.00	6.69	6.59	6.52	5.73	4.30	4.30
4.00	3.30	3.62	3.42	2.80	2.16	2.33

Tables IIA and IIB reveal that method I is quite stable in detecting the shift in terms of the ARL_1 value (because the prespecified type II error possibility is 0.2, the nominal value should be $1/0.8=1.25$), whereas it seems that method II increases its ARL value when k becomes larger. The reason is that method II uses approximate sample size calculated under the assumption of the small difference in hazard functions under H_1.

4.1.2. Data with censoring. In the second scenario, it is assumed that some failure times are censored in the failure data. The censoring distribution is Exp(0.1) for both in-control and out-of-control data, and there is approximately 10% censoring in the in-control data with this censoring distribution. We need to make clear that the c-chart, the exponential CUSUM chart, and the EWMA chart cannot handle censoring. Therefore, in the setup of these three charts, the censoring events are disregarded. In other words, a censored

Table III. Sample sizes n_2 calculated through methods I and II ($\alpha=0.002$, $\beta=0.20$)

k	p_1 = 0.7 I	0.7 II	0.8 I	0.8 II	0.9 I	0.9 II
1.5	128	118	116	106	107	97
2	44	38	41	35	38	32
2.5	26	21	24	19	22	18
3	18	14	17	13	16	12
4	12	8	11	8	11	7

Qual. Reliab. Engng. Int. **2012**, 28 321–333

Table IVA. ARL_1 of four types of control charts (Weibull data with exponential censoring)

	Rank test chart		*c*-chart		Rank test chart $\times n_2$		Exponential	Exponential
k	I	II	I	II	I	II	CUSUM	EWMA
1.50	1.28	1.38	6.31	7.85	148.48	146.28	37104.45	16392.67
2.00	1.26	1.46	5.89	8.63	51.66	51.1	12451.63	3191.81
2.50	1.25	1.56	5.55	9.94	30.01	29.64	3041.65	856.70
3.00	1.24	1.65	5.41	9.65	21.08	21.45	946.00	382.97
4.00	1.26	1.92	5.12	7.27	13.86	15.36	153.17	106.39

Table IVB. ATS of four types of control charts (Weibull data with exponential censoring)

	Rank test chart		*c*-chart		Exponential	Exponential
k	I	II	I	II	CUSUM	EWMA
1.50	107.59	106.01	495.36	563.36	28839.84	12741.42
2.00	32.39	32.14	143.25	179.42	8301.32	2127.70
2.50	16.83	16.59	71.37	101.23	1802.30	507.23
3.00	10.79	10.98	45.32	61.81	509.39	205.95
4.00	6.14	6.82	24.27	25.13	70.62	49.02

failure time and an adjacent uncensored failure time need to be merged together to obtain a time interval for the *c*-chart, CUSUM, and EWMA charts.

Given β=0.20, the sample sizes of monitoring group n_2 for p_1=0.7, 0.8, and 0.9 calculated through Equations (8) and (9) are listed in Table III. As shown in Table III, most of the sample sizes are a little greater than the corresponding values in Table I because of the censoring.

We still use the sample sizes of p_1=0.8 to construct the rank test chart. The study follows the same procedure as in the case without censoring. The in-control data are assumed to follow the Weibull distribution $W(1, 2)$. As stated earlier, the censoring is assumed to follow the exponential distribution, and in this section, we only study the Weibull data. The final results are listed in Tables IVA and IVB. It can be seen that the conclusions regarding the performance of the control charts in the case without censoring still hold here. Tables IVA and IVB reveal that the ARL_1 and the ATS values for the rank test chart are similar to the values in Tables IIA(a) and IIB (a), which means that the rank test chart works very reliably in both cases: uncensored and censored data.

The values labeled as "rank test chart $\times n_2$" in the table represent the numbers of the failure times, whereas the ARL_1 values of the CUSUM and EWMA control charts are the numbers of "merged" time intervals. Considering the fact that there could be more failure times than merged time intervals in the data with censoring, the values labeled as rank test chart $\times n_2$ will become even smaller if we only count the number of merged time intervals there, that is, the rank test chart is much better than the exponential CUSUM and EWMA control charts. The same conclusion holds by looking at the ATS values listed in Table IVB.

4.2. Evaluation of accuracy of sample size formulae

In this section, we investigate the accuracy of the derived sample size formulae and provide some insights on the influential factors of the accuracy. It is assumed that the in-control data follows $W(4\times 10^{-4}, 2)$ and no censoring exists in the data. We assume that the value of hazard ratio k changes from 1.1 to 4; that is, the scale parameter of out-of-control data is in the range of 4.4 to 16×10^{-4}. Given α and β, once sample sizes are calculated via one of the two methods, two groups consisting of n_1 and n_2 in-control and out-of-control failure times are generated and the log-rank test at an α significance level is conducted to calculate the Z value. This study is repeated 10,000 times, and the empirical type II error probability, $\hat{\beta}$, is the proportion of samples not rejected by the test. A good sample size formula should have an empirical power to the projected one, $1-\beta$. In this study, the values of p_1 are 0.6, 0.7, 0.8, and 0.9. The results of $\hat{\beta}$ for the log-rank test at a 1% significance level are summarized in Table V.

As shown in Table V, in most cases, both sample size formulae have empirical power values close to the projected values. However, the difference between the empirical and the projected power tends to become larger when k increases. This trend is apparent especially when k is larger than 2.5 (the sample sizes now are quite small). For example, when α=0.01, k=4, and p_1=0.6, the sample sizes by method II are n_1=12 and n_2=8. Another reason is that method II gives smaller sample sizes compared with method I (n_1=20 and n_2=13) under the same conditions and thus will have larger $\hat{\beta}$ values. We may also notice that when p_1 is small, the method I gives more accurate results, whereas the method II is better when p_1 is large. The reason is that method II is developed compared with method I in the case of the small difference between the hazard functions. In general, the first method tends to give smaller $\hat{\beta}$ than the projected one. As stated, the second method is used in the case that the difference of hazard functions is small.

Table V. Empirical type II error probability $\hat{\beta}$ of the log-rank test with α=0.01

		k									
p_1	Method	1.1	1.2	1.3	1.5	1.8	2	2.5	3	3.5	4
(a) β=0.1											
0.6	I	0.109	0.107	0.113	0.108	0.108	0.107	0.101	0.094	0.089	0.097
	II	0.111	0.119	0.131	0.153	0.175	0.184	0.210	0.249	0.308	0.338
0.7	I	0.100	0.099	0.097	0.090	0.084	0.078	0.071	0.059	0.053	0.050
	II	0.106	0.114	0.119	0.126	0.145	0.160	0.174	0.195	0.202	0.240
0.8	I	0.094	0.091	0.088	0.081	0.066	0.055	0.049	0.030	0.029	0.028
	II	0.106	0.110	0.107	0.112	0.129	0.134	0.148	0.144	0.170	0.191
0.9	I	0.089	0.080	0.074	0.060	0.052	0.041	0.027	0.017	0.011	0.008
	II	0.102	0.105	0.103	0.108	0.097	0.109	0.110	0.103	0.127	0.112
(b) β=0.2											
0.6	I	0.202	0.211	0.216	0.225	0.233	0.228	0.233	0.244	0.249	0.259
	II	0.215	0.238	0.238	0.268	0.287	0.316	0.342	0.386	0.432	0.421
0.7	I	0.211	0.209	0.204	0.202	0.204	0.190	0.192	0.191	0.192	0.173
	II	0.212	0.224	0.227	0.251	0.261	0.283	0.292	0.337	0.370	0.380
0.8	I	0.207	0.187	0.192	0.183	0.163	0.154	0.145	0.143	0.120	0.129
	II	0.212	0.208	0.213	0.220	0.225	0.241	0.260	0.275	0.282	0.271
0.9	I	0.191	0.181	0.177	0.159	0.142	0.126	0.113	0.083	0.090	0.068
	II	0.192	0.204	0.202	0.200	0.202	0.204	0.223	0.193	0.246	0.246

4.3. Case study

In this study, we use a case study to show the application of a rank test control chart. This data set contains the failure times (censored or uncensored) from multiple CT machines of the same type. Total 686 failure times (664 uncensored and 22 censored) are collected under normal working conditions.

To calculate the sample sizes, we need the survival functions for the uncensored and censored failure times (as shown in Equation (5) and (8)). We can fit the Weibull distribution to the data set and estimate the parameters for the uncensored intervals as a=0.6942 and b=0.5556. For censored intervals, we have the two parameters as a=0.9626 and b=0.5102. If the hazard ratio we would like to monitor is k=2, we can calculate the sample sizes with preset values of p_1=0.85, p_2=0.15, α=0.01 and β=0.20 by using Equation (9), and the values are n_1=222 and n_2=40. Notice that the censoring distribution is assumed to be unchanged under the alternative when we calculate the sample size.

The one-sided rank test chart is illustrated in Figure 5. There are 26 points in the chart, and each subgroup contains both uncensored and censored failure times. The value of UCL is 2.326, and we can see that the process is out of control at the 12th plot, which means the hazard function of the 12th subgroup has significantly deviated from the baseline hazard function. In addition, it shows that the machine has an increased hazard rate after the 12th point because most of the points are above the center line. This increase in the hazard rate could be caused by changes in usage pattern or environmental change of the CT machines.

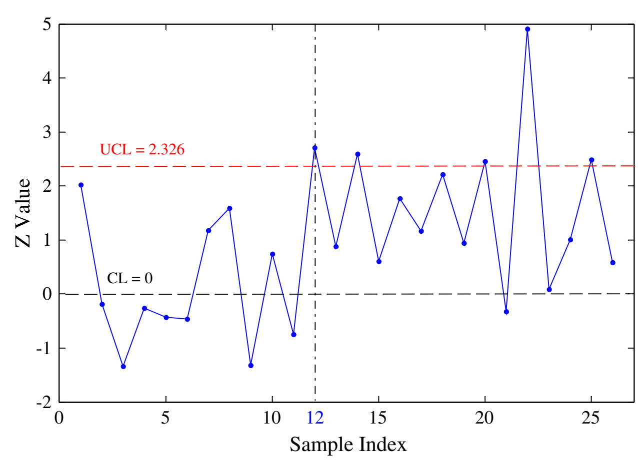

Figure 5. Rank test chart for the computerized tomography machines

5. Concluding remarks

Taking advantage of the weighted rank tests, this article develops a control charting procedure for the hazard rate of failure events extracted from the event logs. The generic formulae for the OC functions are derived to show the relationship between type I error probability, type II error probability, sample size, and hazard rate change. Because the log-rank test is widely used in biomedical and reliability analysis to compare hazard rates or survival curves, we put more emphasis on the control chart design on the basis of the log-rank test. In numerical studies, the effectiveness of the proposed monitoring procedure is illustrated and compared with some current techniques, such as the *c*-chart, the exponential CUSUM chart, and the exponential EWMA chart. Furthermore, the accuracy of sample size formulae is investigated and found satisfactory. The results show that, compared with the other three charts, the rank test chart has significantly better performance for nonexponential data, such as the Weibull data, but worse performance for exponential data.

It needs to be pointed out that the monitoring scheme developed in this article is based on a two-sample test. However, if the hazard function of the system under normal conditions is known, that is, there is an external estimate of h_1 (t), we can turn to the nonparametric one-sample test. The two-sample method can be revised to handle the one-sample test. In addition, although the proposed method can detect nonproportional shift, we mainly focused on the detection of a step proportional hazard shift, which is a relatively simple but common shift model. It is interesting to investigate the performance of the chart under other specific shift models. One important step will be the selection of the weight function in the rank test statistic. The general theory on how to derive an optimal test for the two-sample test is developed by Andersen *et al*[21].

There are some disadvantages when using the rank test control chart in practice. First, because the chart does not require a specific distributional assumption, it could perform worse when the data follows well a specific distribution required by other methods. Second, to calculate the required sample sizes, the data distributions (including failure and censoring events) need to be estimated from the historical in-control data. The calculation itself is complicated by involving the integrations of weight functions, hazard functions, and survival functions for both in-control and out-of-control processes. Third, in practice, it may be difficult for the practitioners specify the intended alternative hazard rate function h_2 (t) to monitor according to the practical requirements. How to overcome the disadvantages is worth studying, and the results along this direction will be presented in the future.

Appendix

Proof of the result regarding the OC function in Equation (5)

We consider the general random censorship model. Suppose the survival times $X_{j1}, X_{j2}, \ldots, X_{j,n_j}$ are independent and identically distributed. with survival function S_j and hazard function h_j, j=1, 2. In the data, we observe (T_{ji}, δ_{ji}), i=1,2,..., n_j} with $T_{ji}=\min(X_{ji}, C_{ji})$ and $\delta_{ji}=I(X_{ji}\leq C_{ji})$, where C_{ji} is the censoring time, and we assume that $C_{j1}, C_{j2}, \ldots, C_{j,n_j}$ are i.i.d. with the survival function $\Pr(C_{ji}>t)=G_j(t)$ and are independent of the survival times.

According to Andersen *et al.*[21] we have $Y_j^{(n)}(s)/n \xrightarrow{P} y_j(s) = p_j S_j(s) G_j(s-)$ as $n\to\infty$, j=1, 2, for all $s\in[0, \tau]$, where τ={t: Pr(min(T_1, T_2, C_1, C_2)>t)>0}. That is, $Y_j^{(n)}(s)/n$ and $Y_\bullet^{(n)}(s)/n$ uniformly converge to $y_j(s)=p_jS_j(s)G_j(s-)$ and $y_\bullet(s) = \sum_{j=1}^{2} y_j(s)$ because $Y_\bullet^{(n)}(s)/n = \sum_{j=1}^{2} Y_j^{(n)}(s)$. According to Lemma V.2.1,[21(p360)] for the Tarone–Ware Family and the Harrington–Fleming Family of test statistics Z_j, j =1, 2 as defined in Equation (4), $\frac{1}{\sqrt{n}}(Z_1, Z_2)^{\mathrm{T}}$ is asymptotically $N(\mathbf{0},\boldsymbol{\Sigma})$, with

$$\sigma_{hj} = \int_0^\infty w^2(s)\frac{y_h(s)}{y_\bullet(s)}\left(\delta_{hj} - \frac{y_j(s)}{y_\bullet(s)}\right)h(s)y_\bullet(s)ds \tag{A1}$$

where δ_{hj} is an indicator variable taking 1 if h=j and 0 otherwise.

On the other hand, under the alternative H_1, $\frac{1}{\sqrt{n}}(Z_1, Z_2)^{\mathrm{T}}$ is approximately $\boldsymbol{N}(\boldsymbol{\zeta},\boldsymbol{\Gamma})$, where $\boldsymbol{\zeta}=(\zeta_1, \zeta_2)$ given by[22]

$$\zeta_j = \sqrt{n}\cdot \int_0^\infty w(s)\sum_{r=1}^{2} y_r(s)\left(\delta_{jr} - \frac{y_j(s)}{y_\bullet(s)}\right)h_r(s)ds, j = 1, 2, \text{and} \tag{A2}$$

$$\Gamma_{hj} = \int_0^\infty w^2(s)\left[\sum_{r=1}^{2} y_r(s)\left(\delta_{hr} - \frac{y_h(s)}{y_\bullet(s)}\right)\left(\delta_{jr} - \frac{y_j(s)}{y_\bullet(s)}\right)h_r(s)\right]ds, h, j = 1, 2 \tag{A3}$$

Because the null hypothesis will be rejected when $|Z_2|\geq Z_{\alpha/2}$ at an α significance level, given the type II error probability β, it can be shown that

$$\frac{1}{\sqrt{n}}|\zeta_2(t)| = \left(Z_{\alpha/2}\sqrt{\frac{\sigma_{22}}{n}} + Z_\beta\sqrt{\frac{\Gamma_{22}}{n}}\right) \tag{A4}$$

where |•| is the absolute value. Inserting $\xi_2(t)$, σ_{22}, and Γ_{22}, after some simplifications, we can obtain Equation (5).

Acknowledgment

The authors thank the editor and the referees for their valuable comments and suggestions.

References

1. Andrews JD, Moss TR. *Reliability and Risk Assessment* (2nd edn). ASME Press: New York, 2002.
2. Grigg O, Farewell V. An Overview of Risk-Adjusted Charts. *Journal of the Royal Statistical Society, Series A* 2004; **167**:523–539.
3. Woodall WH. The Use of Control Charts in Health-Care and Public-Health Surveillance. *Journal of Quality Technology* 2006; **38**:89–104.
4. Steiner SH, Cook RJ, Farewell VT. Risk-Adjusted Monitoring of Binary Surgical Outcomes. *Medical Decision Making* 2001; **21**:163–169.
5. Sego LH, Reynolds MR, Woodall WH. Risk-adjusted Monitoring of Survival Times. *Statistics in Medicine* 2009; **28**:1386–1401.
6. Lucas JM. Counted Data CUSUM's. *Technometrics* 1985; **27**:129–144.
7. Borror CM, Champ CW, Rigdon SE. Poisson EWMA Control Charts. *Journal of Quality Technology* 1998; **30**:352–361.
8. Vardeman S, Ray D. Average Run Lengths for CUSUM Schemes when Observations Are Exponentially Distributed. *Technometrics* 1985; **27**:145–150.
9. Gan FF. Design of Optimal Exponential CUSUM Control Charts. *Journal of Quality Technology* 1994; **26**:109–124.
10. Gan FF. Designs of One- and Two-Sided Exponential EWMA Charts. *Journal of Quality Technology* 1998; **30**:55–69.
11. Nelson LS. A Control Chart for Parts-Per-Million Nonconforming Items. *Journal of Quality Technology* 1994; **30**:239–240.
12. Steiner SH, MacKay RJ. Monitoring Processes with Highly Censored Data. *Journal of Quality Technology* 2000; **32**:199–208.
13. Steiner SH, MacKay RJ. Monitoring Processes with Data Censored Owing to Competing Risks by Using Exponentially Weighted Moving Average Control Charts. *Journal of the Royal Statistical Society Series C- Applied Statistics* 2001; **50**:293–302.
14. Harrington DP, Fleming TR. A Class of Rank Test Procedures for Censored Survival Data. *Biometrika* 1982; **69**:553–566.
15. Yang Z, Xie M, Kuralmani V, Tsui K-L. On the Performance of Geometric-charts with Estimated Parameters. *Journal of Quality Technology* 2002; **34**:448–458.
16. Klein JP, Moeschberger ML. *Survival Analysis: Techniques for Censored and Truncated Data*. Springer-Verlag: New York, 2003.
17. Meeker WQ, Escobar LA. *Statistical Methods for Reliability Data*. John Wiley & Sons: New York, 1998.
18. Tarone RE, Ware J. On distribution-free tests for equality of survival distributions. *Biometrika* 1997; **64**:156–160.
19. Wasserman GS. R*eliability Verification, Testing, and Analysis in Engineering Design*. Marcel Dekker: New York, 2003.
20. Montgomery DC. *Introduction to Statistical Quality Control* (5th edn). John Wiley & Sons: New York, 2005.
21. Andersen PK, Borgan O, Gill RD, Keiding N. *Statistical Models Based on Counting Processes*. Springer: New York, 1993.
22. Jung S-H, Hui S. Sample Size Calculation for Rank Tests Comparing *K* Survival Distributions. *Lifetime Data Analysis* 2002; **8**:361–373.