Process Monitoring ROC Curve for Evaluating Dynamic Screening Methods

Peihua Qiu[1], Zhiming Xia[2] and Lu You[1]

[1]Department of Biostatistics, University of Florida, Gainesville, Florida 32610, U.S.A.

[2]School of Mathematics, Northwest University, Xi'an, Shaanxi 710027, P.R. China

Abstract

In practice, we often need to sequentially monitor the performance of individual subjects or processes, so that interventions can be made in a timely manner to avoid unpleasant consequences (e.g., strokes or airplane crashes) once the longitudinal patterns of their performance variables deviate significantly from the regular patterns of well-functioning subjects or processes. Some statistical methods are available to handle this dynamic screening (DS) problem. Because the performance of the DS methods is related to their signal times, the conventional false positive rate (FPR) and false negative rate (FNR) cannot be effective in measuring their performance. So far, there is no existing metrics in the literature for properly measureing the performance of DS methods. In this paper, we aim to fill this gap by proposing a new performance evaluation approach, called process monitoring ROC curve, which properly combines the signal times with (FPR,FNR). Numerical examples and theoretical justifications show that this approach provides an effective tool for measuring the performance of DS methods.

Key Words: Dynamic screening; Longitudinal data; Online monitoring; Performance measure; Sequential diagnostics; Statistical process control.

1 Introduction

The SHARe Framingham Heart Study of the National Heart, Lung and Blood Institute (cf., Cupples et al. 2007) recruited many residents at Framingham MA to the study. Each recruited person was followed a number of times, and her/his medical indices, including the total cholesterol level (mg/100ml), blood pressures, smoking status, and more, were recorded during each followup. One main purpose of the study was to identify major risk factors of cardiovascular diseases (CVDs), including stroke. Such risk factors are called *performance variables* in this paper. After the major performance variables are identified, the next important question is how we can identify people with irregular longitudinal patterns of the performance variables as early as possible, so that interventions

can be made in a timely manner to avoid any unpleasant consequences (e.g., stroke). Besides the applications for disease early detection mentioned above, this dynamic screening (DS) problem has many other applications. For instance, airplanes, cars, and other durable products need to be checked regularly or occasionally about their performance variables. If the observations of the performance variables of a given product are significantly worse than the values of a typical well-functioning product of the same age, then some proper adjustments or interventions are desirable.

In the literature, there are some existing methods that are relevant to the DS problem. One group of methods is in the area of *longitudinal data analysis (LDA)*. By an LDA method, we can construct confidence intervals for the mean values of the performance variables at different time points based on an observed dataset of some well-functioning subjects, called an in-control (IC) dataset hereafter. Then, a new subject can be identified as the one who has an irregular longitudinal pattern if his/her observations of the performance variables fall outside the confidence intervals. Some existing methods for constructing such confidence intervals include Chen and Jin (2005), Li (2011), Liang and Zeger (1986), Lin and Carroll (2000), Ma et al. (2012), Wang (2003), Xiang et al (2013), and Zhao and Wu (2008). However, the confidence interval approach is not sequential and thus may not be efficient in handling the DS problem, mainly because it does not make use of all history data of a subject when making a decision about his/her performance at the current time point. For instance, when we are interested in monitoring a patient's total cholesterol level over time, if her/his total cholesterol levels are consistently above the mean total cholesterol levels of healthy people for a long time period, then that patient should be identified as one who has an irregular total cholesterol level pattern, even if the observed total cholesterol level at any given time point is within the related confidence interval. Obviously, the confidence interval approach cannot achieve that goal. Another group of methods relevant to the DS problem belongs to the research area of *statistical process control (SPC)* (cf., Hawkins and Olwell 1998, Qiu 2014). By a SPC control chart, we can monitor each subject sequentially, and a signal could be given once the chart detects a shift in the longitudinal pattern of the performance variables from an IC status to an out-of-control (OC) status (i.e., the monitored pattern deviates significantly from the regular pattern). However, a conventional SPC chart cannot be applied to the DS problem directly because it requires the assumption that the distribution of the performance variables is unchanged over time when their longitudinal pattern is IC. In most applications of the DS problem, however, such distributions would change over time (e.g., the mean total cholesterol level of healthy people would change

with age). To overcome this limitation, Qiu and Xiang (2014,2015) recently proposed a method, called the dynamic screening system (DySS), for monitoring the longitudinal pattern of a subject sequentially while allowing the IC distribution of the performance variables varying over time. By the DySS method, a person will receive a signal once DySS detects a significant difference by the current observation time between the person's longitudinal pattern of the performance variables and the regular longitudinal pattern of a typical non-diseased person. For modified versions of the DySS method, see papers such as Li and Qiu (2016, 2017), Qiu et al. (2018) and You and Qiu (2019). This method has been shown effective in solving the DS problem.

To evaluate the performance of the methods for solving the DS problem, the conventional false positive rate (FPR), false negative rate (FNR) and the related receiver operating characteristic (ROC) curve approaches (e.g., Obuchowski 2003, Pepe 2003, Qiu and Le 2001, Zhou et al. 2002) may not be good enough because these metrics/methods only consider the proportions of correct or incorrect decisions regarding whether or not a diseased (healthy) person receives a signal from a DS method, but they do not take into account the signal time when the abnormality in the longitudinal pattern of the performance variables is first detected. As a demonstration, assume that there are two DS methods, both of them can perfectly identify all people with a specific disease, but the first one can detect the disease at quite early times while the second one detects the disease quite late. Obviously, the first one performs better than the second one in this rather extreme case. But, such a difference cannot be distinguished by the conventional FPR and FNR. In the literature, another related approach is the time-dependent ROC approach proposed for handling cases when disease status of a subject changes over time in a given time interval (e.g., Heagerty et al. 2000, Heagerty and Zheng 2005, Saha and Heagerty 2010). However, this approach is not for handling sequential monitoring problems. Instead, it is discussed mainly for cases when the disease status of each subject is known at any time point in an interval, and the diagnostic marker is either a time-independent covariate or a time-dependent covariate that its value at any time point is completely specified. In such cases, an ROC curve can be obtained at each time point in the interval. In the current DS problem, we sequentially monitor each person by taking observations of her/his performance variables at a sequence of observation time points, and the observation time points of different people can be all different. We usually make judgments whether the longitudinal pattern of a person's performance variables is IC or not at the observation times only. Therefore, the number of people who have observations available at a given observation time point could be too

small to construct a meaningful ROC curve. For instance, in the stroke data example discussed in Section 4, there is usually 0 or 1 stroke patient who has observation at a specific time point. Also, the time-dependent ROC curves are mainly for evaluating a diagnostic method at different time points. It is not obvious how to use them for evaluating the *overall* performance of a diagnostic method. In the SPC literature, the performance of a sequential monitoring chart is often measured by the IC and OC average run lengths (ARLs), which take into account the signal time of the chart. These metrics are appropriate to use in the conventional SPC problem, in which only one sequential process is under monitoring. In the current DS problem, if the longitudinal pattern of the performance variables of a subject is considered a sequential process, then many sequential processes are involved because a DS method would be used for monitoring the performance variables of many subjects. In such cases, the conventional IC and OC ARLs cannot reflect the overall performance of the DS method regarding its false positive and false negative rates. Therefore, there is no appropriate metrics in the literature so far for measuring the performance of DS methods. This paper aims to fill this gap by proposing a new method for evaluating the overall performance of the DS methods. This method, called the process monitoring ROC (PM-ROC) curve, combines the signal times with the conventional (FPR,FNR) metrics in a reasonable way. Thus, it takes into account both the accuracy and the speed of the decisions made by the DS methods.

The remaining part of the paper is organized as follows. Our proposed method is described in detail in Section 2. Some of its statistical properties are discussed in Section 3. Some simulation examples regarding its performance are presented in Section 4. Then, this method is applied to a stroke dataset from the Framingham heart study in Section 5. Several remarks conclude the article in Section 6.

2 Process Monitoring ROC Curve

2.1 The DS problem

For simplicity, let us assume that there is only one performance variable y involved in the DS problem, and an IC dataset contains m subjects whose observations of y follow the model

$$y_i(t_{ij}^*) = \mu_0(t_{ij}^*) + \varepsilon_i(t_{ij}^*), \qquad t_{ij}^* \in [0,1], \text{ for } j = 1,2,\ldots,J_i^*, i = 1,2,\ldots,m, \tag{1}$$

where t_{ij}^* is the jth observation time of the ith subject, $y_i(t_{ij}^*)$ is the observed value of y at t_{ij}^*, $\mu_0(t_{ij}^*)$ is the mean of $y_i(t_{ij}^*)$, and $\varepsilon_i(t_{ij}^*)$ is the error term with mean 0 and variance $\sigma_y^2(t_{ij}^*)$. At each $t \in [0,1]$, the error term is assumed to have the expression $\varepsilon(t) = \varepsilon_0(t) + \varepsilon_1(t)$, where $\varepsilon_0(t)$ is a random process with mean 0 and covariance function $E[\varepsilon_0(s)\varepsilon_0(t)] = V(s,t)$, for $s,t \in [0,1]$, denoting the impact of all covariates that may affect the response but are not included in the model, $\varepsilon_0(t)$ denotes a zero-mean pure measurement error satisfying the condition that $\varepsilon_0(s)$ and $\varepsilon_0(t)$ are independent when $s \neq t \in [0,1]$, and $\varepsilon_0(t)$ and $\varepsilon_1(t)$ are independent. Without imposing any parametric forms on $\mu_0(t)$, $V(s,t)$ and the distribution of $\varepsilon(t)$, Qiu and Xiang (2014) suggested nonparametric estimates of $\mu_0(t)$ and $V(s,t)$, denoted as $\widehat{\mu}_0(t)$ and $\widehat{V}(s,t)$. From $\widehat{V}(s,t)$, we can define the estimate of $\sigma_y^2(t)$ to be $\widehat{\sigma}_y^2(t) = \widehat{V}(t,t)$. The expressions of $\widehat{\mu}_0(t)$ and $\widehat{V}(s,t)$ are given in an appendix.

Now, for a new subject to monitor, assume that its observations of $y(t)$ are obtained at times $t_1, t_2, \ldots, t_J$ in $[0,1]$. Then, we can monitor its standardized observations

$$\widehat{\varepsilon}(t_j) = \frac{y(t_j) - \widehat{\mu}_0(t_j)}{\widehat{\sigma}_y(t_j)}, \qquad j = 1, 2, \ldots, J. \tag{2}$$

If the new subject is IC, then $\widehat{\varepsilon}(t_j)$ would have an asymptotic mean of 0 and an asymptotic variance of 1. To sequentially monitor the sequence $\{y(t_1), y(t_2), \ldots, y(t_J)\}$ to see whether it follows the regular longitudinal pattern described by (1), we can monitor its standardized version $\{\widehat{\varepsilon}(t_1), \widehat{\varepsilon}(t_2), \ldots, \widehat{\varepsilon}(t_J)\}$ to see whether the mean of each $\widehat{\varepsilon}(t_j)$ is close to 0 and the variance is close to 1. To this end, we can consider using the cumulative sum (CUSUM) chart with the charting statistic

$$C_j^+ = \max\left(0, C_{j-1}^+ + \widehat{\varepsilon}(t_j) - k\right), \qquad j = 1, 2, \ldots, J, \tag{3}$$

where $C_0^+ = 0$, and $k > 0$ is the allowance constant. The chart signals an upward mean shift at t_j if

$$C_j^+ > \rho, \tag{4}$$

where $\rho > 0$ is a control limit. If a downward or two-sided mean shift is a concern in a specific application, then the corresponding CUSUM chart can be constructed accordingly. See the related discussion in Chapter 4 of Qiu (2014) for details. The sequential monitoring scheme (1)–(4) is the DySS procedure discussed in Qiu and Xiang (2014). Its multivariate version is discussed in Qiu and Xiang (2015). In (3)–(4), the conventional CUSUM chart is used. In cases when the data distribution is believed to be quite different from a normal distribution, a nonparametric control

chart can also be used (Qiu 2018). See some related discussion and a numerical example in Section 3.

Note that construction of the DySS procedure (1)–(4) has not taken into account the data correlation. When the data correlation is assumed to follow an AR(1) model, Qiu and Xiang (2014) modified the CUSUM statistic C_j^+ accordingly. In cases when such parametric time series models are unavailable or invalid, we suggested choosing the control limit ρ in (4) by a block bootstrap procedure from the IC data to reach a pre-specified IC ATS value (see its definition in the next subsection). Li and Qiu (2016) proposed an alternative approach to handle correlated data in a DySS system. Since construction of an appropriate control chart is not the focus of the current paper, interested readers are referred to Qiu and Xiang (2014) and Li and Qiu (2016) for more detailed discussions.

2.2 Performance measure of the DS methods

The DySS procedure (1)–(4) is just one example of the existing statistical methods for handling the DS problem, as discussed in Section 1. However, all these methods share a similar structure that a time-dependent statistic (e.g., C_j^+ in (3)) is compared to a threshold value to determine whether a given subject is IC or not at a specific time point. In this part, we describe our proposed method for evaluating the performance of DS methods. For convenience, our description is made for the DySS procedure only, but the new method can actually measure the performance of other DS methods.

To evaluate the performance of the DySS procedure (1)–(4), let D denote the true performance status of a given subject in the entire monitoring time period $[0, 1]$, with "$D = 0$" denoting IC in $[0, 1]$ and "$D = 1$" denoting OC in $[0, 1]$. If we define

$$Y_1^* = \max_{j=1,2,...,J} C_j^+, \tag{5}$$

then the true positive rate (TPR) (or, sensitivity), and the false positive rate (FPR) (or, 1-specificity) of the DySS procedure are defined by

$$\text{FPR}(\rho) = \mathbb{P}\left(Y_1^* > \rho | D = 0\right), \tag{6}$$

$$\text{TPR}(\rho) = \mathbb{P}\left(Y_1^* > \rho | D = 1\right). \tag{7}$$

The ROC curve is generated by the pair $(\mathrm{FPR}(\rho), \mathrm{TPR}(\rho))$ when ρ changes in $(0, \infty)$ (cf., Pepe 2003). The ROC curve provides a general tool for comparing the overall performance of two diagnostic methods. If the ROC curve of one method is completely above the ROC curve of the other method, then we conclude that the former gives a more accurate diagnosis than the latter. If the two curves cross each other, then we can use their summary statistics for comparison purposes. One commonly used summary statistic is the area under the curve (AUC) defined as

$$\mathrm{AUC} = \int_0^1 \mathrm{ROC}(t)\mathrm{d}t, \tag{8}$$

where $\mathrm{ROC}(t) = \mathrm{TPR}(\mathrm{FPR}^{-1}(t))$. By this summary statistic, the larger its value, the better.

The conventional ROC curve evaluates the overall false positive and false negative rates of a diagnosis method. In the current DS problem, however, diagnosis is made in a sequential manner, and signal times of irregular longitudinal pattern should be taken into account when evaluating the performance of a DS method, which cannot be achieved by the conventional ROC curve. In the SPC literature, the average time to signal (ATS) is often used to measure the performance of a sequential monitoring procedure (Qiu 2014). Let

$$\mathrm{T}(\rho) = \begin{cases} \min\{t_j : C_j^+ > \rho\}, & \text{if } Y_1^* > \rho \\ 1, & \text{otherwise} \end{cases}$$

be the signal time of the DS method with the control limit ρ. Then, the IC and OC ATS values are defined, respectively, by

$$\mathrm{ATS}_0(\rho) = \mathbb{E}\left(\mathrm{T}(\rho)|D=0\right), \tag{9}$$

$$\mathrm{ATS}_1(\rho) = \mathbb{E}\left(\mathrm{T}(\rho)|D=1\right). \tag{10}$$

It's easy to see that $\mathrm{ATS}_0(\rho)$ and $\mathrm{ATS}_1(\rho)$ are both non-decreasing functions of ρ. So, we can define

$$\begin{aligned} \overline{T}_{0,r} &= \lim_{\rho\to\infty} \mathrm{ATS}_0(\rho), \quad \overline{T}_{0,l} = \lim_{\rho\to 0} \mathrm{ATS}_0(\rho), \\ \overline{T}_{1,r} &= \lim_{\rho\to\infty} \mathrm{ATS}_1(\rho), \quad \overline{T}_{1,l} = \lim_{\rho\to 0} \mathrm{ATS}_1(\rho). \end{aligned} \tag{11}$$

It can be checked that $\overline{T}_{0,r} = \overline{T}_{1,r} = 1$, but $\overline{T}_{0,l}$ and $\overline{T}_{1,l}$ could take any values in $[0, 1]$ because C_j^+ could be 0 at any time points (cf., its definition in (3)). Ideally, we hope that the value of $\mathrm{ATS}_1(\rho)$ can be small and the value of $\mathrm{ATS}_0(\rho)$ can be large. But, for a given ρ, it is easy to check that $\mathrm{ATS}_1(\rho)$ will be larger if $\mathrm{ATS}_0(\rho)$ is set larger, which is similar to the relationship between the type-I and type-II error probabilities in the hypothesis testing setup. So, in practice, we often set the value of $\mathrm{ATS}_0(\rho)$ at a given level, and then make $\mathrm{ATS}_1(\rho)$ as small as possible.

The metrics $\mathrm{ATS}_0(\rho)$ and $\mathrm{ATS}_1(\rho)$ measure the signal times of a DS method. Their values, however, cannot reflect the overall false positive and false negative rates of the DS method. As a demonstration, assume that there are 50 non-diseased people and 50 diseased people involved in a study. One DS method gives signals to all 50 non-diseased people at quite late time points, and does not signal for the 50 diseased people. Another DS method gives signals to 25 non-diseased people at quite early time points, and does not give signals to the remaining 25 non-diseased people and all 50 diseased people. In such cases, the second method is better than the first method in terms of the false positive and false negative rates. But, it is the other way around in terms of $\mathrm{ATS}_0(\rho)$ and $\mathrm{ATS}_1(\rho)$. Therefore, both the conventional ROC curve and the conventional metrics of $\mathrm{ATS}_0(\rho)$ and $\mathrm{ATS}_1(\rho)$ cannot properly describe the performance of DS methods.

In this paper, we aim to develop a new performance evaluation system which takes into account both the signal times and the conventional false positive and false negative rates in a reasonable way. To this end, let us define the *dynamic false positive rate (DFPR)* (or, *dynamic 1-specificity*) to be

$$\mathrm{DFPR}(\rho) = \left[1 - \frac{\mathrm{ATS}_0(\rho) - \overline{T}_{0,l}}{1 - \overline{T}_{0,l}}\right] \times \mathrm{FPR}(\rho), \tag{12}$$

and the *dynamic true positive rate (DTPR)* (or, *dynamic sensitivity*) to be

$$\mathrm{DTPR}(\rho) = \left[1 - \frac{\mathrm{ATS}_1(\rho) - \overline{T}_{1,l}}{1 - \overline{T}_{1,l}}\right] \times \mathrm{TPR}(\rho). \tag{13}$$

In (12) and (13), both quantities $1-(\mathrm{ATS}_0(\rho)-\overline{T}_{0,l})/(\overline{T}_{1,r}-\overline{T}_{1,l})$ and $1-(\mathrm{ATS}_1(\rho)-\overline{T}_{1,l})/(\overline{T}_{1,r}-\overline{T}_{1,l})$ decrease from 1 to 0 when ρ increases from 0 to ∞. So, $\mathrm{DFPR}(\rho)$ and $\mathrm{DTPR}(\rho)$ have the following properties:

(i) Both $\mathrm{DFPR}(\rho)$ and $\mathrm{DTPR}(\rho)$ change from 1 to 0 when ρ increases from 0 to ∞,

(ii) $\mathrm{DFPR}(\rho)$ is smaller if $\mathrm{ATS}_0(\rho)$ is larger or $\mathrm{FPR}(\rho)$ is smaller, and

(iii) $\mathrm{DTPR}(\rho)$ is larger if $\mathrm{ATS}_1(\rho)$ is smaller or $\mathrm{TPR}(\rho)$ is larger.

The property (i) guarantees that the pair ($\mathrm{DFPR}(\rho)$, $\mathrm{DTPR}(\rho)$) changes from (0,0) to (1,1) when ρ changes in $[0,\infty)$. The resulting curve $\{(\mathrm{DFPR}(\rho), \mathrm{DTPR}(\rho)), \rho \in [0,\infty)\}$ is called *process monitoring ROC (PM-ROC)* curve hereafter. The properties (ii) and (iii) make the quantities $\mathrm{DFPR}(\rho)$ and $\mathrm{DTPR}(\rho)$ legitimate metrics for measuring the longitudinal performance of the DS

method, because its performance would be better when $\text{ATS}_0(\rho)$ and/or $\text{TPR}(\rho)$ are larger or when $\text{FPR}(\rho)$ and/or $\text{ATS}_1(\rho)$ are smaller, which is intuitively reasonable.

The random variable Y_1^* in (5) provides us useful information about the overall evidence in the observed data of a subject under monitoring regarding whether or not s/he is IC in $[0, 1]$. Therefore, Y_1^* is closely related to the conventional $\text{FPR}(\rho)$ and $\text{TPR}(\rho)$ defined in (6) and (7) and to the conventional ROC curve as well. By Theorem 2 in Section 3 below, we can find another random variable Y_2^* that is independent of Y_1^* and has the following properties:

$$\mathbb{P}(Y_2^* \le \rho | D = 0) = \frac{\text{ATS}_0(\rho) - \overline{T}_{0,l}}{1 - \overline{T}_{0,l}}, \tag{14}$$

$$\mathbb{P}(Y_2^* \le \rho | D = 1) = \frac{\text{ATS}_1(\rho) - \overline{T}_{1,l}}{1 - \overline{T}_{1,l}}, \tag{15}$$

where $\overline{T}_{0,l}$ and $\overline{T}_{1,l}$ are defined in (11). From (14) and (15), it can be seen that Y_2^* is related to the signal time. Now, let us define

$$Y^* = \min(Y_1^*, Y_2^*). \tag{16}$$

Then, we have

$$\begin{aligned} \mathbb{P}(Y^* > \rho | D = 0) &= \mathbb{P}(Y_1^* > \rho, Y_2^* > \rho | D = 0) \\ &= \mathbb{P}(Y_1^* > \rho | D = 0)\mathbb{P}(Y_2^* > \rho | D = 0) = \text{DFPR}(\rho), \end{aligned} \tag{17}$$

$$\begin{aligned} \mathbb{P}(Y^* > \rho | D = 1) &= \mathbb{P}(Y_1^* > \rho, Y_2^* > \rho | D = 1) \\ &= \mathbb{P}(Y_1^* > \rho | D = 1)\mathbb{P}(Y_2^* > \rho | D = 1) = \text{DTPR}(\rho). \end{aligned} \tag{18}$$

Therefore, from (16), the random variable Y^* combines information about both the signal time (through Y_2^*) and the false positive and true positive rates (through Y_1^*). From (18), $\text{DTPR}(\rho)$ can be interpreted as the conditional probability that the combined information of a shift is above the control limit ρ for all diseased people. Similarly, from (17), $\text{DFPR}(\rho)$ can be interpreted as the conditional probability for all non-diseased people.

2.3 Estimation of the PM-ROC curve

To compute $\text{DFPR}(\rho)$ and $\text{DTPR}(\rho)$ defined in (12) and (13), assume that there are $n_{\overline{D}}$ non-diseased people (group 1) and n_D diseased people (group 2) in a training dataset. Observations of these people follow the following model

$$y_i(t_{ij}) = \mu_i(t_{ij}) + \varepsilon_i(t_{ij}), \;\; t_{ij} \in [0, 1], \; i = 1, 2, \ldots, n, \; j = 1, 2, \ldots, J, \tag{19}$$

where $n = n_{\overline{D}} + n_D$, $\mu_i(t) = \mu_0(t)$ for $i = 1, 2, \ldots, n_{\overline{D}}$, $\mu_i(t) = \mu_0(t) + \delta_i I_{\{t>\tau_i\}}$ for $i = n_{\overline{D}} + 1, n_{\overline{D}} + 2, \ldots, n$, $\tau_i > 0$ is a mean shift time, and $\delta_i \neq 0$ is the mean shift size. Let us further assume that the DySS procedure (1)–(4) with the control limit ρ detects $n_1(\rho)$ people in group 1 and $n_2(\rho)$ people in group 2. Then, the conventional FPR(ρ) is $n_1(\rho)/n_{\overline{D}}$, and the conventional TPR(ρ) is $n_2(\rho)/n_D$. The quantities $\widehat{\text{ATS}}_0(\rho)$ and $\widehat{\text{ATS}}_1(\rho)$ are calculated by averaging the signal times of the DySS procedure in groups 1 and 2, respectively. For the DySS procedure, $\overline{T}_{0,r} = \overline{T}_{1,r} = 1$, and $\overline{T}_{0,l}$ and $\overline{T}_{1,l}$ can be estimated by $\text{ATS}_0(0)$ and $\text{ATS}_1(0)$, respectively, which can be computed easily from the training dataset. To estimate the PM-ROC curve, we need to determine a sequence of ρ values. To this end, let the values of the charting statistic in (3) of people in the training data be $\{C_{ij}^+, i = 1, 2, \ldots, n, j = 1, 2, \ldots, J\}$. Then, the ρ values can be chosen to be the order statistics of $\{C_{ij}^+, i = 1, 2, \ldots, n, j = 1, 2, \ldots, J\}$.

For the PM-ROC curve, we can construct its pointwise $100(1-\alpha)\%$ confidence interval by the bootstrap resampling method as follows. First, randomly select $n_{\overline{D}}$ non-diseased people with replacement from group 1 of the training data and n_D diseased people from group 2 of the training data. Observations of the selected $n = n_{\overline{D}} + n_D$ people constitute a bootstrap sample. Second, at each ρ value, DFPR(ρ) and DTPR(ρ) can be estimated from the bootstrap sample. Third, the first two steps are repeated for B times, and B sets of estimates of $(\text{DFPR}(\rho), \text{DTPR}(\rho))$ can be obtained. Then, the α_1-th and $(1-\alpha_2)$-th quantiles of the B estimates of DFPR(ρ) form the $100(1-\alpha)\%$ confidence interval for DFPR(ρ), where $\alpha_1 + \alpha_2 = \alpha$. The $100(1-\alpha)\%$ confidence interval for DTPR(ρ) can be constructed similarly. To construct the $100(1-\alpha)\%$ confidence interval for $(\text{DFPR}(\rho), \text{DTPR}(\rho))$, for each estimate of $(\text{DFPR}(\rho), \text{DTPR}(\rho))$ from a bootstrap sample, compute its Euclidean distance from $(\widehat{\text{DFPR}}(\rho), \widehat{\text{DTPR}}(\rho))$ which are obtained from the original training data. Then, the B Euclidean distances obtained from the B bootstrap estimates of $(\text{DFPR}(\rho), \text{DTPR}(\rho))$ can be ordered. The bootstrap estimates corresponding to the α_1-th and $(1-\alpha_2)$-th quantiles of the B Euclidean distances form a $100(1-\alpha)\%$ confidence interval for $(\text{DFPR}(\rho), \text{DTPR}(\rho))$. In practice, we can simply choose $\alpha_1 = \alpha_2 = \alpha/2$.

2.4 Some statistical properties

In this part, we discuss some statistical properties of the estimated PM-ROC curve. The consistency of the estimated PM-ROC curve is established in Theorem 1. Existence of a random variable Y_2^* so that (14) and (15) are valid is confirmed in Theorem 2. Finally, area under the

PM-ROC curve is discussed in Theorem 3.

Theorem 1. *In model (19), assume that (i) the observation times* $\{t_{ij}, i = 1, 2, \ldots, n, j = 1, \cdots, J\}$ *for each person follows a distribution with a continuous density function* $f(t)$ *in* $[0, 1]$*, (ii) observed data of different people are independent, and (iii) both the* n_D *diseased people and the* $n_{\overline{D}}$ *non-diseased people are simple random samples from the population of diseased people and the population of non-diseased people, respectively. Then, we have*

$$\widehat{ATS_0}(\rho) \xrightarrow{p} ATS_0(\rho), \ \ \widehat{FPR}(\rho) \xrightarrow{p} FPR(\rho), \ \ \textit{for each } \rho \in R^+, \ \ \textit{as } n_{\overline{D}} \to \infty; \tag{20}$$

$$\widehat{ATS_1}(\rho) \xrightarrow{p} ATS_1(\rho), \ \ \widehat{TPR}(\rho) \xrightarrow{p} TPR(\rho), \ \ \textit{for each } \rho \in R^+, \ \ \textit{as } n_D \to \infty; \tag{21}$$

$$\widehat{PM\text{-}ROC}(\rho) \xrightarrow{p} PM\text{-}ROC(\rho), \ \ \textit{for each } \rho \in R^+, \ \ \textit{as } n_{\overline{D}}, n_D \to \infty. \tag{22}$$

Theorem 2. *There exists a random variable* Y_2^* *that is independent of* Y_1^* *in (5) such that (14) and (15) are valid.*

Similar to AUC defined in (8), we can define the area under the PM-ROC curve as

$$\text{DAUC} = \int_0^1 \text{DTPR}\left(\text{DFPR}^{-1}(t)\right) \text{d}t.$$

For this quantity, we have the following result.

Theorem 3. *Let* $Y_D^* = (Y^*|D = 1)$ *and* $Y_{\overline{D}}^* = (Y^*|D = 0)$*, where* Y^* *is defined in (16). Then, we have*

$$DAUC = \mathbb{P}(Y_{\overline{D}}^* \leq Y_D^*). \tag{23}$$

By (23), the quantity DAUC can be interpreted as the probability of the event $\{Y_{\overline{D}}^* \leq Y_D^*\}$. The proofs of Theorems 1-3 are given in the appendices.

3 Simulation Study

In this part, we present some simulation results about the proposed PM-ROC approach for measuring the performance of DS methods. First, let us consider cases when $\mu_0(t) = 1 + 0.3t^{1/2}$, $\delta_i = \delta$ and $\tau_i = 0$, for $i = n_{\overline{D}} + 1, n_{\overline{D}} + 2, \ldots, n$, and $J = 10$ in model (19). In that model, the random errors $\{\varepsilon_i(t_{ij})\}$ are generated independently from the distribution $N(0, [\mu_0(t_{ij}) + \delta_i]^2)$, and

$\{t_{ij}, j = 1, 2, \ldots, 10\}$ are equally spaced in $[0, 1]$ for each i. In this setup, the control limit ρ of the CUSUM chart (3)-(4) is computed based on 1,000 replicated simulations as follows. For given values of k and ρ, the CUSUM chart is applied to observed data of 1,000 non-diseased people, and FPR can be estimated by the proportion of non-diseased people who get signals from the chart. Then, for the given value of k and a given level of FPR, ρ can be determined by searching its value such that the given level of FPR is reached. In cases when the allowance constant k takes values in $\{0, 0.1, 0.2, 0.5, 0.75, 1.0\}$ and FPR takes values in $\{0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6\}$, the computed ρ values are presented in Table 1. From the table, it can be seen that the computed ρ value is smaller when FPR is larger or k is larger.

Table 1: Computed control limit ρ of the CUSUM chart (3)-(4) for some k and FPR values.

k	FPR=0.01	0.05	0.1	0.2	0.3	0.4	0.5	0.6
0	8.321	6.269	5.175	4.193	3.448	2.924	2.491	2.169
0.1	7.337	5.571	4.526	3.601	2.966	2.496	2.134	1.851
0.2	6.972	4.824	3.903	3.091	2.615	2.163	1.893	1.617
0.5	5.110	3.181	2.620	2.031	1.690	1.460	1.231	1.033
0.75	3.108	2.295	1.923	1.459	1.187	1.006	0.820	0.657
1.0	2.416	1.751	1.454	1.071	0.849	0.692	0.517	0.356

Now, let us consider an example with the following two cases

Case I: There are 100 diseased people and 100 healthy people. The mean shift size δ changes from 1.5 to 2.5 with a step of 0.1,

Case II: There are 100 diseased people and 100 healthy people. The mean shift size for each diseased person is two times the one considered in Case I.

In the CUSUM chart (3)-(4), let us choose $k = 1$ and FPR = 0.01. In such a case, $\rho = 2.416$ by Table 1. Then, the computed values of DTPR, TPR and ATS_1 of the chart in the two cases are presented in Figure 1(a). From the plot, it can be seen that the conventional TPR is about the same in Cases I and II when δ is relative large (e.g., $\delta > 2.1$). So, this criterion cannot distinguish the different performance of the CUSUM chart in the two cases. As a comparison, the chart performs much better in Case II by the criterion of ATS_1, which is intuitively reasonable because the shift

size in Case II is larger than the one in Case I. Our proposed criterion DTPR can also distinguish the performance of the chart in the two cases well.

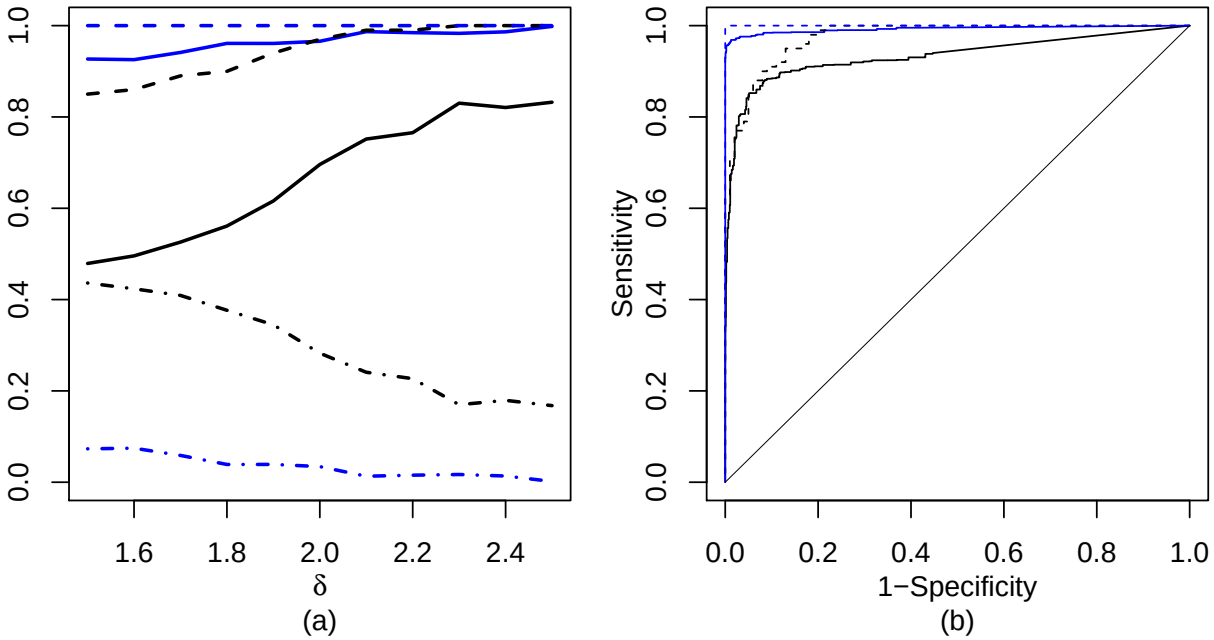

Figure 1: (a) Computed values of DTPR (solid line), TPR (dashed line), and ATS_1 (dot-dashed line) of the CUSUM chart (3)-(4) with $k = 1$ and FPR $= 0.01$ in Cases I (black color) and II (blue color). (b) Computed PM-ROC curves (solid lines) and ROC curves (dashed lines) of the chart (3)-(4) with $k = 1$ and its control limit ρ varying in Cases I (black color) and II (blue color) when δ is fixed at 1.5 (i.e., the shift size is 1.5 in Case I, and 3.0 in Case II).

Now, let us fix the value of δ to be 1.5 (i.e., the shift size is 1.5 in Case I, and 3.0 in Case II). In the CUSUM chart (3)-(4), let us still choose $k = 1$ and let its control limit ρ vary. In such cases, the PM-ROC curves and the ROC curves of the chart are shown in Figure 1(b). From the plot, it can be seen that the ROC curves in the two cases cannot be distinguished when 1-Specificity (i.e., FPR) is above 0.3. The two PM-ROC curves, however, are quite different in the entire range of 1-Specificity.

Next, we consider another example with the following two cases:

Case III: There are 100 diseased people and 100 healthy people. The mean shift size δ changes from 3 to 4 with a step of 0.1,

Case IV: There are 100 diseased people and 100 healthy people. For the first 90 diseased people,

the shift size is 2.5δ, where δ is the shift size considered in Case III. For the remaining 10 patients, the shift size is 0.25δ.

The results in this example are shown in Figure 2, in the same setup as that in Figure 1. From Figure 2(a), it can be seen that the ATS_1 values are about the same in Cases III and IV when δ is in $[3, 3.6]$. Thus, the criterion ATS_1 cannot distinguish the performance of the chart in the two cases when $\delta \in [3, 3.6]$. As a comparison, the chart performs better in Case III than its performance in Case IV by the conventional TPR criterion, which is intuitively reasonable because most of the 10 diseased people with very small shift sizes in Case IV would not receive signals and almost all other diseased people in the two cases would receive signals. Our proposed criterion DTPR can distinguish the different performance of the chart in the two cases well in this example. Figure 2(b) shows the PM-ROC curves and ROC curves of the CUSUM chart (3)-(4) when δ is fixed at 3.0, $k = 1$ and its control limit ρ changes. This plot shows that both the PM-ROC curves and the ROC curves can distinguish the performance of the chart in the two cases reasonably well, but the PM-ROC curves can better distinguish the two cases when 1-Specificity is above 0.3.

As a summary, the above two examples demonstrate the limitations of the conventional performance criteria TPR (i.e., sensitivity) and ATS_1 in measuring the performance of DS methods. While the criterion TPR does not take into account the signal times of the DS methods, the criterion ATS_1 cannot reflect the overall true positive rate. So, there are cases when either one of them cannot distinguish the different performance of the DS methods. As a comparison, our proposed criterion DTPR can overcome their limitations and distinguish the different performance of the DS methods in various different cases well.

Next, we study the consistency of the proposed PM-ROC curve. In model (19), assume that $\mu_i(t) = \mu_0(t) + \delta_i I_{\{t>\tau_i\}}$, $\mu_0(t) = 1 + 0.3t^{1/2}$, $n = 100, 500$ or 1000, $\{t_{ij}, i = 1, 2, \ldots, n, j = 0, \ldots, J\}$ are generated as in the previous examples, and the random errors are generated i.i.d. from $N(0, 0.5^2)$. In the CUSUM chart (3)-(4), assume that we choose $k = 1$ and $\text{FPR} = 0.1$. So, by Table 1, $\rho = 1.454$. Because the PM-ROC curve is based on (DFPR(ρ),DTPR(ρ)), which in turn depend on $\text{ATS}_0(\rho), \text{FPR}(\rho), \text{ATS}_1(\rho)$, and $\text{TPR}(\rho)$, we study the consistency of their estimates $\widehat{\text{ATS}_0}(\rho), \widehat{\text{FPR}}(\rho), \widehat{\text{ATS}_1}(\rho)$, and $\widehat{\text{TPR}}(\rho)$ next. Let us consider the following six cases regarding how the mean shift times and sizes $\{(\tau_i, \delta_i), i = 1, 2, \ldots, n\}$ are generated:

Case (a): IC case with $\delta_i = 0$ for all i.

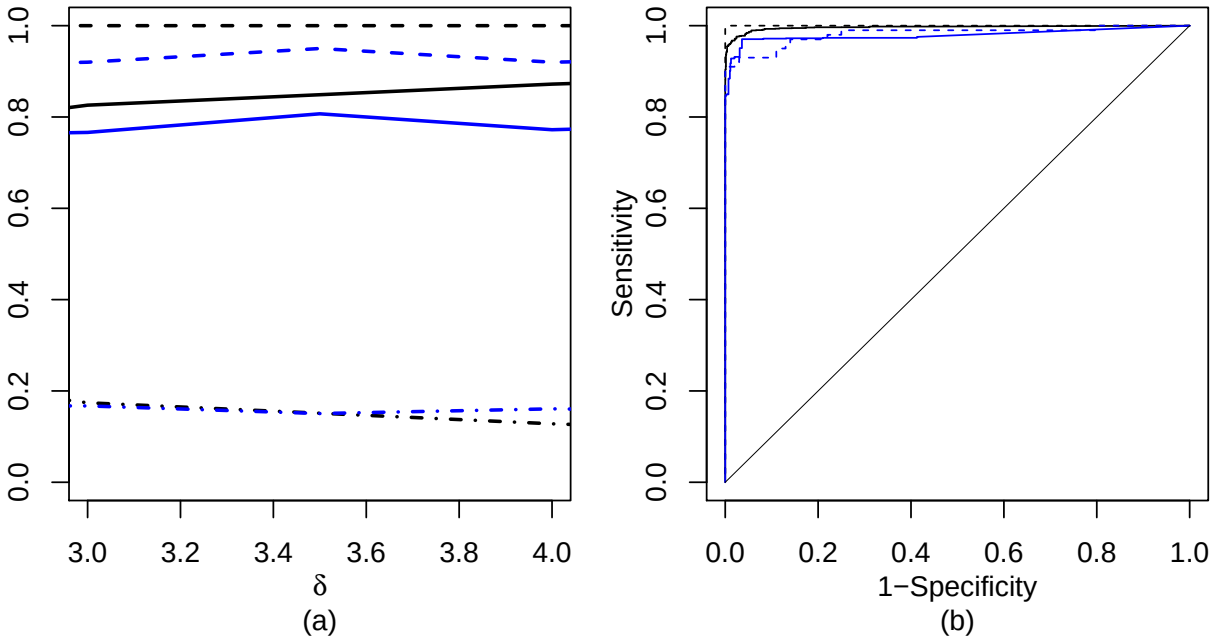

Figure 2: (a) Computed values of DTPR (solid line), TPR (dashed line), and ATS_1 (dot-dashed line) of the CUSUM chart (3)-(4) with $k = 1$ and FPR $= 0.01$ in Cases III (black color) and IV (blue color). (b) Computed PM-ROC curves (solid lines) and ROC curves (dashed lines) of the chart (3)-(4) with $k = 1$ and its control limit ρ varying in Cases III (black color) and IV (blue color) when δ is fixed at 3.

Case (b): Case when $\delta_i = 0.1$ and $\tau_i = 0.3$ for all i.

Case (c): Case when $\{\delta_i, i = 1, 2, \ldots, n\}$ are generated i.i.d. from the distribution $N(1, 0.2^2)$ and $\tau_i = 0.3$ for all i.

Case (d): Case when $\{\delta_i, i = 1, 2, \ldots, n\}$ are generated i.i.d. from the distribution $N(1, 0.2^2)$ and $\{\tau_i, i = 1, 2, \ldots, n\}$ are generated i.i.d. from the distribution $N(0.3, 0.2^2)$.

Case (e): Case when $\{\delta_i = 0, i = 1, 2, \ldots, n/4\}$ and $\{\delta_i = 1, \tau_i = 0.3, i = n/4 + 1, \ldots, n\}$.

Case (f): Case when $\{\delta_i = 0, i = 1, 2, \ldots, 3n/4\}$ and $\{\delta_i = 1, \tau_i = 0.3, i = 3n/4 + 1, \ldots, n\}$.

Obviously, all people in Case (a) are IC and all people in Cases (b)-(d) are OC with either deterministic or random mean shift times and sizes. Cases (e) and (f) consider two cases when some people are IC and some other people are OC, with Case (f) having more IC people than Case (e). In Case (a), we focus on the performance of $\widehat{\text{ATS}_0}(\rho)$ and $\widehat{\text{FPR}}(\rho)$, in Cases (b)-(d) we focus on

the performance of $\widehat{\text{ATS}_1}(\rho)$ and $\widehat{\text{TPR}}(\rho)$, and in Cases (e) and (f) we focus on the performance of $\widehat{\text{ATS}_0}(\rho)$, $\widehat{\text{ATS}_1}(\rho)$ and $\widehat{\text{TPR}}(\rho)$. The results based on 1,000 replicated simulations are summarized in Table 2 below. From the table, it can be seen that: i) $\widehat{\text{ATS}_0}(\rho)$ is larger and larger and $\widehat{\text{FPR}}(\rho)$ is closer and closer to FPR = 0.1 in Case (a) when n increases, ii) $\widehat{\text{TPR}}(\rho)$ is larger and larger in Cases (b)-(d) when n increases, (iii) $\widehat{\text{ATS}_1}(\rho)$ is close to the true value of $\tau_i = 0.3$ in Cases (b) and (c), and close to the mean 0.3 of $\{\tau_i, i = 1, 2, \ldots, n\}$ in Case (d), (iv) $\widehat{\text{TPR}}(\rho)$ is larger when n gets larger in Cases (e) and (f), (v) $\widehat{\text{ATS}_0}(\rho)$ is larger and $\widehat{\text{ATS}_1}(\rho)$ is smaller in Case (e) than those in Case (f). Result (v) can be explained by the fact that there are more OC people and less IC people in case (e), compared to Case (f). Because some IC people will receive signals from the control chart and these signals could be much later than 0.3, the values of $\widehat{\text{ATS}_1}(\rho)$ in Cases (e) and (f) are generally larger than those in Cases (b)-(d). All other results are consistent with the theoretical results discussed in Section 3.

Table 2: Calculated values of $\widehat{\text{ATS}_0}(\rho)$, $\widehat{\text{FPR}}(\rho)$, $\widehat{\text{ATS}_1}(\rho)$, and $\widehat{\text{TPR}}(\rho)$ in cases when $k = 1$ and FPR = 0.1 ($\rho = 1.454$ in such cases). Numbers in parentheses are standard errors.

Case	Estimate	$n = 100$	500	1000
(a)	$\widehat{\text{ATS}_0}(\rho)$	0.950(0.015)	0.952(0.007)	0.953(0.005)
	$\widehat{\text{FPR}}(\rho)$	0.110(0.030)	0.098(0.013)	0.099(0.009)
(b)	$\widehat{\text{ATS}_1}(\rho)$	0.352(0.015)	0.350(0.007)	0.348(0.005)
	$\widehat{\text{TPR}}(\rho)$	0.990(0.010)	0.996(0.003)	0.999(0.001)
(c)	$\widehat{\text{ATS}_1}(\rho)$	0.347(0.015)	0.345(0.007)	0.351(0.005)
	$\widehat{\text{TPR}}(\rho)$	0.990(0.010)	0.994(0.004)	0.995(0.002)
(d)	$\widehat{\text{ATS}_1}(\rho)$	0.304(0.019)	0.302(0.008)	0.303(0.006)
	$\widehat{\text{TPR}}(\rho)$	0.990(0.010)	0.992(0.004)	0.995(0.002)
(e)	$\widehat{\text{ATS}_0}(\rho)$	0.978(0.022)	0.930(0.019)	0.955(0.011)
	$\widehat{\text{ATS}_1}(\rho)$	0.476(0.018)	0.484(0.009)	0.481(0.006)
	$\widehat{\text{TPR}}(\rho)$	0.987(0.013)	0.989(0.005)	0.991(0.004)
(f)	$\widehat{\text{ATS}_0}(\rho)$	0.946(0.023)	0.935(0.010)	0.948(0.007)
	$\widehat{\text{ATS}_1}(\rho)$	0.501(0.023)	0.485(0.014)	0.496(0.011)
	$\widehat{\text{TPR}}(\rho)$	0.920(0.055)	0.989(0.020)	0.992(0.006)

In all previous examples, the conventional CUSUM chart (3)-(4) is used. Next, we consider

a nonparametric CUSUM chart suggested by Li et al. (2010). To use that method, a reference sample is needed. To this end, in each simulation, standardized observations of $n_R = 100$ extra IC subjects are generated as in Case (f) of the example of Table 2, and they are used as the reference sample and denoted as $R = \{\tilde{\varepsilon}_k : k = 1, \ldots, Jn_R\}$. Then, for the jth standardized observation of the new subject to monitor (cf., its definition in (2)), let W_j be the rank of $\widehat{\varepsilon}(t_j)$ in the combined set of $\widehat{\varepsilon}(t_j)$ and R. The nonparametric CUSUM charting statistic for detecting an upward mean shift is defined by

$$\tilde{C}_j^+ = \max\left(0, \tilde{C}_{j-1}^+ + (W_j - \mu_{W_j})/\sigma_{W_j} - \tilde{k}\right),$$

where $\tilde{C}_0^+ = 0$, $\tilde{k} > 0$ is an allowance constant, $\mu_{W_j} = (Jn_R + 1)/2$, and $\sigma_{W_j}^2 = Jn_R(Jn_R + 2)/12$. Then, we consider scenarios when the total sample size $n_{\bar{D}} + n_D = 500$ and the ratio $n_{\bar{D}} : n_D$ is 1:1, 10:1, 20:1 or 30:1. In these four scenarios, $(n_{\bar{D}}, n_D)$=(250,250), (455,45), (476,24) and (484,16), respectively. Under the same simulation setup as that in Table 2, the calculated values of $\widehat{\text{ATS}_0}(\rho)$, $\widehat{\text{ATS}_1}(\rho)$, and $\widehat{\text{TPR}}(\rho)$ are shown in Table 3. From the table, it can be seen that when the ratio $n_{\bar{D}} : n_D$ increases, the calculated values of $\widehat{\text{ATS}_0}(\rho)$, $\widehat{\text{ATS}_1}(\rho)$ and $\widehat{\text{TPR}}(\rho)$ are quite stable, but the standard errors of $\widehat{\text{ATS}_1}(\rho)$ and $\widehat{\text{TPR}}(\rho)$ get larger, which is intuitively reasonable because the sample size n_D of the OC subjects gets smaller in such cases.

Table 3: Calculated values of $\widehat{\text{ATS}_0}(\rho)$, $\widehat{\text{ATS}_1}(\rho)$, and $\widehat{\text{TPR}}(\rho)$ when $n_D + n_{\bar{D}} = 500$, the ratio $n_{\bar{D}} : n_D$ changes among 1:1, 10:1, 20:1 and 30:1, and other setups are the same as those in Case (f) of Table 2. Numbers in parentheses are standard errors.

	$n_{\bar{D}} : n_D = 1:1$	10:1	20:1	30:1
$\widehat{\text{ATS}_0}(\rho)$	0.946(0.016)	0.947(0.013)	0.947(0.013)	0.946(0.013)
$\widehat{\text{ATS}_1}(\rho)$	0.534(0.015)	0.534(0.027)	0.537(0.037)	0.535(0.046)
$\widehat{\text{TPR}}(\rho)$	0.986(0.008)	0.986(0.018)	0.984(0.025)	0.985(0.030)

4 Application to the Stroke Data

In this section, we apply the proposed new performance measuring method to a dataset obtained from the SHARe Framingham Heart Study mentioned in Section 1. This dataset contains longitudinal observations of 1055 people, 1028 of them did not have any strokes and the other 27 had at least one stroke during the study. Each person was followed 7 times. Observation times

(in years) were different for different people in the range $[9, 85]$. In this study, we use the diastolic blood pressure as the performance variable, and treat all stroke patients as people in the case group and all non-stroke patients as people in the control group. Two conspicuous outliers with diastolic blood pressure being 0 are imputed by the sample median. Following the analysis in Qiu and Xiang (2014) about this dataset, we first compute the point estimates of $\mu_0(t)$ and $\sigma_y(t)$ in model (1) from the data of the control group, and then compute the standardized observations (cf., expression (2)) of the entire dataset, including those in the case and control groups.

The CUSUM chart (3)–(4) is then applied to the standardized observations. For some selections of (k, ρ) in the chart, the computed results of ATS_0, ATS_1, FPR, TPR, DFPR, DTPR, TPR/FPR, DTPR/DFPR, and DTPR/DFPR-TPR/FPR are presented in Table 4. The results presented in the second line of each entry in the table are the 90% confidence intervals (CIs) for the related parameters, computed by the bootstrap procedure described at the end of Section 2.3. From the table, it can be seen that 1) ATS_1 values are uniformly smaller than ATS_0 values, 2) TPR is uniformly larger than FPR, and 3) DTPR is uniformly larger than DFPR. The first conclusion says that the average signal time for people in the case group is earlier than that for people in the control group, which is intuitively reasonable. The second conclusion says that the positive rate in the case group is generally larger than that of the control group, implying that the performance variable is effective in diagnosing the disease. The third conclusion takes into account both the first and second conclusions, and demonstrates the effectiveness of the performance variable better. To better see this, let us look at the last three columns of the table. We can see that the ratio DTPR/DFPR is uniformly larger than the ratio TPR/FPR, with quite large margins in most entries. Consequently, all values of DTPR/DFPR − TPR/FPR shown in the last column of the table are positive. In the stroke data, there are 1028 non-stroke people and only 27 stroke people involved. So, the two groups of data are unbalanced in size. This explains why the CIs for ATS_1 are generally wider than those for ATS_0. The CIs for DTPR/DFPR are generally wider than those for TPR/FPR because the estimates of DTPR/DFPR involve the estimates of ATS_0 and ATS_1, besides the estimates of TPR and FPR. So, the randomness in the estimates of DTPR/DFPR is generally larger than the randomness in the estimates of TPR/FPR. Now, let us take a look at the CIs for DTPR/DFPR − TPR/FPR. It can be seen that all these CIs are completely above 0, implying that DTPR/DFPR is significantly larger than TPR/FPR in all cases considered. To better see this, the estimated value of DTPR/DFPR − TPR/FPR and its 90% confidence intervals

in cases when $k = 0.05, 0.1, 0.2$ or 0.5 and the control limit ρ changes from 0 to 3 are presented in Figure 3. It can be seen from the plots that DTPR/DFPR $-$ TPR/FPR is significantly larger than 0 in all cases, implying the significant benefit to use the proposed performance measures DTPR and DFPR in comparison with the regular TPR and FPR. The related ROC curves and PM-ROC curves in cases when $k = 0.05, 0.1, 0.2$ and 0.5 are shown in Figure 4. From the plots of the figure, it can be seen that most parts of the PM-ROC curves are above the related ROC curves. It should also be pointed out that the results in Figure 4 are consistent with those in Table 4. To better see this, Figure 5 shows the point on the PM-ROC curve (denoted by a dark dot) and the corresponding point on the ROC curve (denoted by a cross) in each case when the control limit ρ changes among 0.5, 1.0, 1.5 or 2.0 (different columns) and when k equals 0.05, 0.1, 0.2 or 0.5 (different rows). The slop of the solid (dotted) line connecting the dot (cross) with the origin is actually DTPR/DFPR (TPR/FPR). So, from the plots in the figure, it can be seen that DTPR/DFPR is always large than TPR/FPR.

5 Concluding Remarks

This paper proposes new metrics for measuring the performance of the DS methods. The new metrics combine the conventional (FPR,FNR) with the signal times of the related monitoring scheme. It has been shown that it provides a more effective evaluation of the DS methods in various cases, compared to the conventional (FPR,FNR) and (ATS_0,ATS_1) metrics. Some issues about the proposed new metrics still need to be addressed in our future research. For instance, we define GFPR(ρ) and GTPR(ρ) in (12) and (13). The definition of these quantities can actually be generalized to

$$\begin{aligned} \text{DFPR}(\rho) &= g\left(1 - \frac{\text{ATS}_0(\rho) - \overline{T}_{0,l}}{\overline{T}_{0,r} - \overline{T}_{0,l}}, \text{FPR}(\rho)\right) \\ \text{DTPR}(\rho) &= g\left(1 - \frac{\text{ATS}_1(\rho) - \overline{T}_{1,l}}{\overline{T}_{1,r} - \overline{T}_{1,l}}, \text{TPR}(\rho)\right), \end{aligned}$$

where $g(x, y)$ is an increasing continuous function of both x and y such that $g(0, 0) = 0$ and $g(1, 1) = 1$. In (12) and (13), $g(x, y)$ is chosen to be xy. It can also be chosen to be $g(x, y) = \alpha x + (1 - \alpha)y$, where $\alpha \in [0, 1]$ is a weight. In our future research, we need to study which specific function for $g(x, y)$ is more reasonable to use in different cases.

Table 4: Results of the proposed new performance measures in the example about the stroke data when k and ρ in the CUSUM chart (3)–(4) take various different values. The numbers in the second line of each entry provide the 90% confidence intervals for the related parameters.

k	ρ	ATS_0	ATS_1	FPR	TPR	DFPR	DTPR	TPR/FPR	DTPR/DFPR	DTPR/DFPR-TPR/FPR
0.05	0.5	0.358	0.215	0.708	0.852	0.572	0.774	1.203	1.352	0.149
		(0.338,0.379)	(0.110,0.321)	(0.686,0.731)	(0.741,0.963)	(0.544,0.602)	(0.632,0.910)	(1.033,1.358)	(1.103,1.593)	(0.036,0.260)
	1.0	0.493	0.258	0.583	0.852	0.372	0.731	1.462	1.967	0.505
		(0.472,0.515)	(0.159,0.362)	(0.558,0.608)	(0.741,0.963)	(0.343,0.401)	(0.592,0.859)	(1.261,1.647)	(1.597,2.329)	(0.302,0.704)
	1.5	0.580	0.295	0.501	0.852	0.265	0.694	1.700	2.618	0.918
		(0.559,0.602)	(0.202,0.392)	(0.476,0.526)	(0.741,0.963)	(0.241,0.290)	(0.564,0.816)	(1.473,1.928)	(2.108,3.158)	(0.614,1.267)
	2.0	0.648	0.335	0.431	0.815	0.191	0.626	1.891	3.283	1.392
		(0.628,0.670)	(0.234,0.442)	(0.405,0.455)	(0.702,0.926)	(0.169,0.211)	(0.483,0.764)	(1.573,2.206)	(2.437,4.182)	(0.847,1.998)
0.1	0.5	0.388	0.223	0.683	0.852	0.534	0.765	1.247	1.432	0.185
		(0.366,0.410)	(0.131,0.333)	(0.659,0.706)	(0.741,0.963)	(0.504,0.565)	(0.624,0.892)	(1.068,1.400)	(1.159,1.678)	(0.056,0.303)
	1.0	0.516	0.260	0.558	0.852	0.346	0.729	1.526	2.110	0.585
		(0.493,0.539)	(0.169,0.366)	(0.532,0.585)	(0.741,0.963)	(0.318,0.373)	(0.598,0.849)	(1.297,1.718)	(1.685,2.490)	(0.354,0.804)
	1.5	0.603	0.297	0.475	0.852	0.241	0.692	1.794	2.875	1.081
		(0.582,0.625)	(0.210,0.397)	(0.448,0.500)	(0.741,0.963)	(0.217,0.265)	(0.564,0.806)	(1.523,2.021)	(2.294,3.433)	(0.726,1.439)
	2.0	0.671	0.346	0.405	0.815	0.170	0.617	2.014	3.623	1.610
		(0.651,0.691)	(0.253,0.449)	(0.380,0.430)	(0.704,0.926)	(0.151,0.189)	(0.475,0.746)	(1.663,2.299)	(2.715,4.525)	(1.001,2.217)
0.2	0.5	0.434	0.223	0.635	0.852	0.482	0.774	1.341	1.605	0.264
		(0.412,0.457)	(0.120,0.335)	(0.610,0.660)	(0.741,0.963)	(0.450,0.513)	(0.613,0.916)	(1.143,1.525)	(1.268,1.908)	(0.094,0.417)
	1.0	0.558	0.282	0.511	0.852	0.303	0.716	1.668	2.364	0.696
		(0.535,0.581)	(0.180,0.387)	(0.484,0.535)	(0.741,0.963)	(0.275,0.330)	(0.575,0.849)	(1.418,1.904)	(1.829,2.870)	(0.398,0.999)
	1.5	0.642	0.329	0.433	0.815	0.208	0.640	1.882	3.079	1.197
		(0.620,0.664)	(0.224,0.435)	(0.406,0.459)	(0.704,0.926)	(0.185,0.231)	(0.496,0.789)	(1.583,2.188)	(2.360,3.880)	(0.729,1.763)
	2.0	0.715	0.376	0.349	0.778	0.133	0.568	2.227	4.260	2.033
		(0.695,0.736)	(0.271,0.486)	(0.325,0.373)	(0.630,0.889)	(0.116,0.150)	(0.410,0.725)	(1.808,2.633)	(3.002,5.681)	(1.134,3.071)
0.5	0.5	0.559	0.273	0.504	0.852	0.351	0.755	1.691	2.149	0.458
		(0.536,0.580)	(0.179,0.370)	(0.481,0.529)	(0.741,0.963)	(0.325,0.381)	(0.616,0.887)	(1.450,1.937)	(1.727,2.585)	(0.224,0.698)
	1.0	0.684	0.382	0.373	0.741	0.186	0.558	1.988	3.002	1.014
		(0.664,0.704)	(0.269,0.497)	(0.350,0.396)	(0.593,0.889)	(0.166,0.207)	(0.400,0.744)	(1.599,2.380)	(2.070,4.096)	(0.442,1.738)
	1.5	0.762	0.408	0.286	0.741	0.108	0.535	2.590	4.969	2.379
		(0.742,0.781)	(0.299,0.519)	(0.265,0.310)	(0.593,0.889)	(0.093,0.125)	(0.381,0.713)	(2.074,3.119)	(3.404,6.848)	(1.299,3.741)
	2.0	0.810	0.497	0.234	0.630	0.070	0.386	2.686	5.483	2.797
		(0.792,0.826)	(0.370,0.620)	(0.214,0.256)	(0.481,0.778)	(0.059,0.083)	(0.229,0.567)	(2.004,3.374)	(3.172,8.360)	(1.138,5.007)

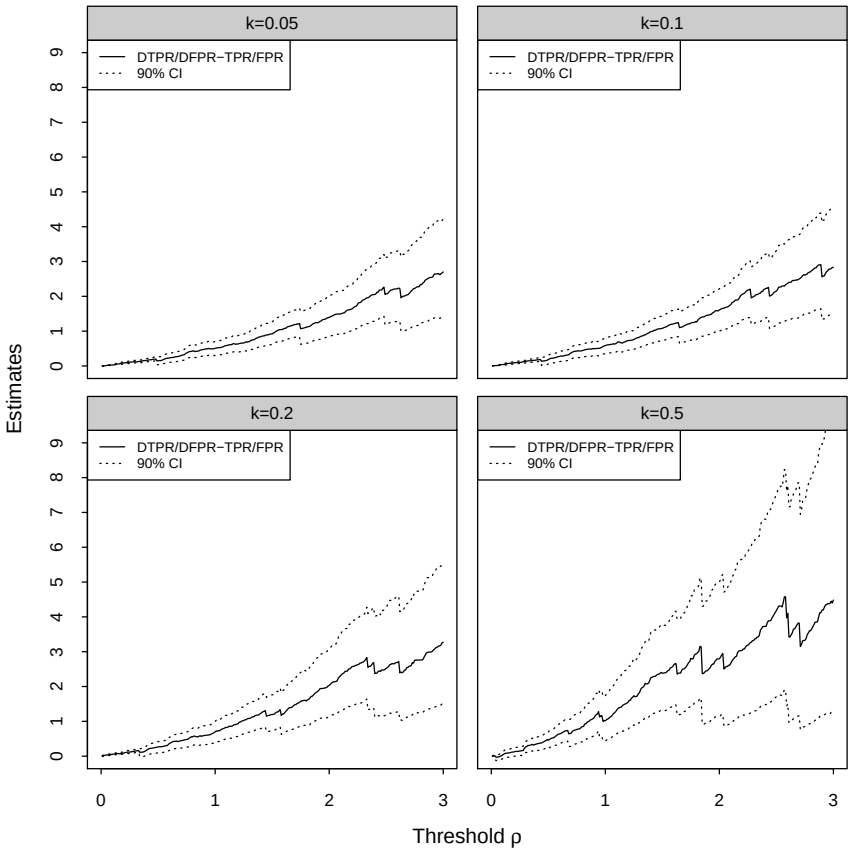

Figure 3: Estimates of DTPR/DFPR − TPR/FPR (solid lines) when $k = 0.05, 0.1, 0.2, 0.5$ and ρ changes from 0 to 3. The 90% bootstrap confidence intervals of DTPR/DFPR − TPR/FPR are shown by dotted lines.

Acknowledgments: The authors thank the editor, the associate editor and two referees for many constructive comments and suggestions, which improved the quality of the paper greatly. This research is supported in part by an NSF grant.

Supplementary Materials

CodesData.zip: This file contains the computer codes of the proposed method and the real data used in the paper.

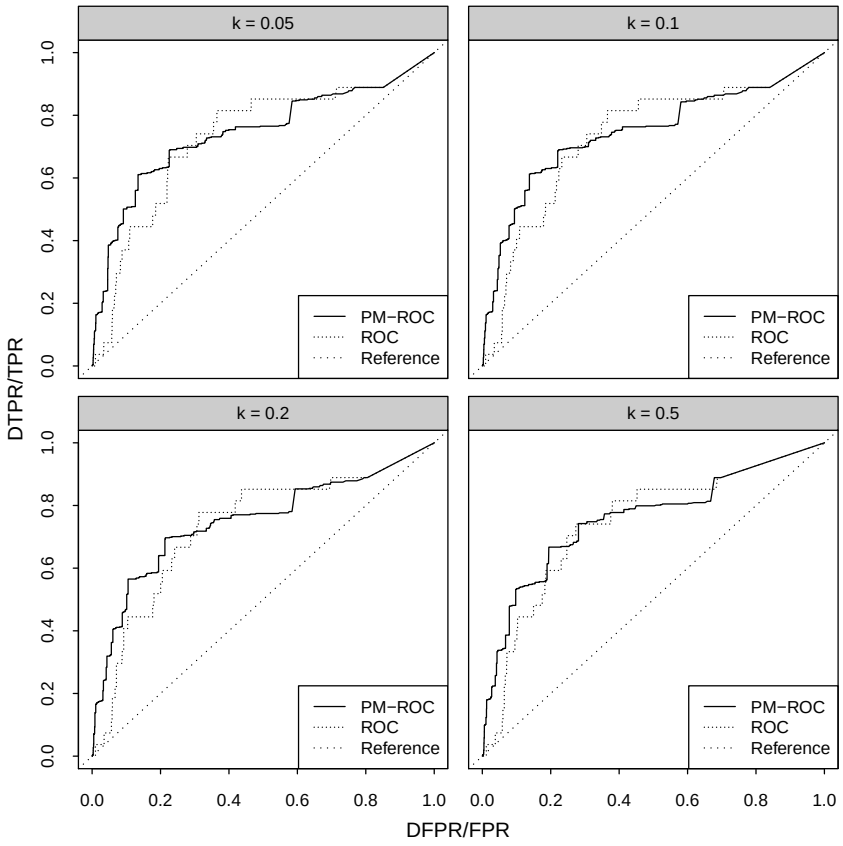

Figure 4: PM-ROC curves (solid lines) and ROC curves (dashed lines) in cases when $k = 0.05, 0.1, 0.2$ and 0.5.

Appendices

A Estimation of $\mu_0(t)$ and $V(s,t)$

First, we obtain the following initial estimate of $\mu_0(t)$ by the regular local linear kernel smoothing procedure:

$$\widetilde{\mu}_0(t) = \mathbf{e}_1' \left(\sum_{i=1}^{m} X_i' K_{ih}(t) X_i \right)^{-1} \Big(\sum_{i=1}^{m} X_i' K_{ih}(t) \mathbf{y}_i \Big),$$

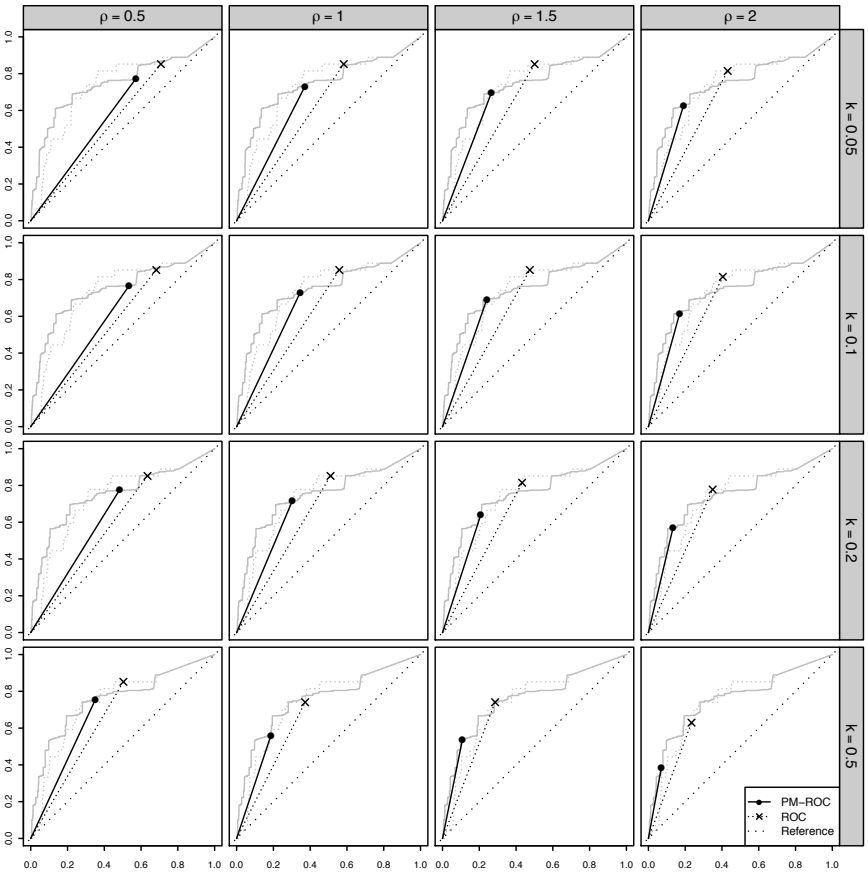

Figure 5: Point on the PM-ROC curve (denoted by a dark dot) and the corresponding point on the ROC curve (denoted by a cross) in each case when ρ changes among 0.5, 1.0, 1.5 or 2.0 and k changes among 0.05, 0.1, 0.2 or 0.5. The gray solid and dotted curves are the PM-ROC and ROC curves.

where $\mathbf{e}_1$ is a 2-dimensional vector with the first element being 1 and the other elements being 0,

$$X_i = \begin{pmatrix} 1 & (t^*_{i1} - t) \\ \vdots & \vdots \\ 1 & (t^*_{iJ^*_i} - t) \end{pmatrix},$$

$$K_{ih}(t) = \mathrm{diag}\left\{K\big((t^*_{i1} - t)/h\big)/h, \ldots, K\big((t^*_{iJ^*_i} - t)/h\big)/h\right\},$$

$K(\cdot)$ is a kernel function, $h > 0$ is a bandwidth, and $\mathbf{y}_i = (y_i(t^*_{i1}), \ldots, y_i(t^*_{iJ^*_i}))'$.

Second, we define residuals

$$\widetilde{\epsilon}_i(t^*_{ij}) = y_i(t^*_{ij}) - \widetilde{\mu}_0(t^*_{ij}), \text{ for } j = 1, \ldots, J^*_i \text{ and } i = 1, \ldots, m,$$

and when $s \neq t$, estimate $V(s,t)$ by

$$\widehat{V}(s,t) = [A_1(s,t)V_{00}(s,t) - A_2(s,t)V_{10}(s,t) - A_3(s,t)V_{01}(s,t)]/\, B(s,t),$$

where $A_1(s,t) = S_{20}(s,t)S_{02}(s,t) - S_{11}^2(s,t)$, $A_2(s,t) = S_{10}(s,t)S_{02}(s,t) - S_{01}(s,t)S_{11}(s,t)$, $A_3(s,t) = S_{01}(s,t)S_{20}(s,t) - S_{10}(s,t)S_{11}(s,t)$, $B(s,t) = A_1(s,t)S_{00}(s,t) - A_2(s,t)S_{10}(s,t) - A_3(s,t)S_{01}(s,t)$,

$$S_{l_1 l_2}(s,t) = \frac{1}{Nh^2}\sum_{i=1}^{m}\sum_{j_1=1}^{J_i^*}\sum_{\substack{j_2=1 \\ j_2\neq j_1}}^{J_i^*}\Big(\frac{t_{ij_1}^* - s}{h}\Big)^{l_1}\Big(\frac{t_{ij_2}^* - t}{h}\Big)^{l_2}K\Big(\frac{t_{ij_1}^* - s}{h}\Big)K\Big(\frac{t_{ij_2}^* - t}{h}\Big),$$

$$V_{l_1 l_2}(s,t) = \frac{1}{Nh^2}\sum_{i=1}^{m}\sum_{j_1=1}^{J_i^*}\sum_{\substack{j_2=1 \\ j_2\neq j_1}}^{J_i^*}\widetilde{\epsilon}_i(t_{ij_1}^*)\widetilde{\epsilon}_i(t_{ij_2}^*)\Big(\frac{t_{ij_1}^* - s}{h}\Big)^{l_1}\Big(\frac{t_{ij_2}^* - t}{h}\Big)^{l_2}K\Big(\frac{t_{ij_1}^* - s}{h}\Big)K\Big(\frac{t_{ij_2}^* - t}{h}\Big),$$

and $N = \sum_{i=1}^{m} J_i^*(J_i^* - 1)$.

Third, estimate $V(t,t)$ by

$$\widehat{V}(t,t) = \mathbf{e}_1'\Big(\sum_{i=1}^{m} X_i' W_i X_i\Big)^{-1}\Big(\sum_{i=1}^{m} X_i' W_i \widetilde{\boldsymbol{\varepsilon}}_i^2\Big),$$

where $\widetilde{\boldsymbol{\varepsilon}}_i^2 = (\widetilde{\epsilon}_i^2(t_{i1}^*), \ldots, \widetilde{\epsilon}_i^2(t_{iJ_i^*}^*))'$, $W_i = K_{ih}^{1/2}(t)[J_i(t)\Sigma_i J_i(t)]^{-1}K_{ih}^{1/2}(t)$, $J_i(t) = \mathrm{diag}\{I_{\{|t_{i1}^* - t| \leq h\}}, \ldots, I_{\{|t_{iJ_i^*}^* - t| \leq h\}}\}$, and $I_{\{\cdot\}}$ is an indicator function that equals 1 when the event is "true" and 0 otherwise.

Fourth, $\mu_0(t)$ is estimated by

$$\widehat{\mu}_0(t) = \mathbf{e}_1'\Big(\sum_{i=1}^{m} X_i' \widehat{W}_i X_i\Big)^{-1}\Big(\sum_{i=1}^{m} X_i' \widehat{W}_i \mathbf{y}_i\Big),$$

where $\widehat{W}_i = K_{ih}^{1/2}(t)[J_i(t)\widehat{\Sigma}_i J_i(t)]^{-1}K_{ih}^{1/2}(t)$, and

$$\widehat{\Sigma}_i = \begin{pmatrix} \widehat{V}(t_{i1}^*, t_{i1}^*) & \cdots & \widehat{V}(t_{i1}^*, t_{iJ_i^*}^*) \\ \vdots & \ddots & \vdots \\ \widehat{V}(t_{iJ_i^*}^*, t_{i1}^*) & \cdots & \widehat{V}(t_{iJ_i^*}^*, t_{iJ_i^*}^*) \end{pmatrix}.$$

B Proof of Theorem 1

Given observations of n_D diseased people $\{y_i(t_{ij}), i = 1, 2, \ldots, n_D, j = 1, 2, \ldots, J\}$, we will show the results in (21) about $\widehat{\mathrm{ATS}_1}(\rho)$ and $\widehat{\mathrm{TPR}}(\rho)$. For the i-th diseased person, we can calculate the

charting statistic values $\{C_{ij}^{+}, i = 1, 2, \ldots, n_D, j = 1, 2, \cdots, J\}$ by (3), and the chart gives a signal at the time

$$T_i(\rho) = \begin{cases} \inf\{t_{ij'} : C_{ij'}^{+} \geq \rho, j' = 1, 2, \cdots, J\}, & \text{if a signal is given,} \\ t_{i,J}, & \text{otherwise,} \end{cases}$$

where $\rho \in R^{+}$ and $i = 1, 2, \cdots, n_D$. Furthermore, we can record whether a signal is actually given by using the indicator variable

$$I_i(\rho) = \begin{cases} 1, & \text{if a signal is given,} \\ 0, & \text{otherwise.} \end{cases}$$

Then, it is obvious that

$$\widehat{\text{ATS}}_1(\rho) = \frac{1}{n_D}\sum_{i=1}^{n_D} T_i(\rho), \tag{24}$$

$$\widehat{\text{TPR}}(\rho) = \frac{1}{n_D}\sum_{i=1}^{n_D} I_i(\rho). \tag{25}$$

By the assumptions given in Theorem 1, for each $\rho \in R^{+}$, $T_i(\rho), i = 1, 2, \ldots, n_D$ are i.i.d.. So, by (24) and the law of large numbers we have $\widehat{\text{ATS}_1}(\rho) \xrightarrow{p} \text{ATS}_1(\rho)$. Similarly, $I_i(\rho), i = 1, 2, \ldots, n_D$ are also i.i.d., and consequently by (25) we have $\widehat{\text{TPR}}(\rho) \xrightarrow{p} \text{TPR}(\rho)$. Therefore, results in (21) are valid. Results in (20) can be verified in a similar way.

To prove the result in (22), first we notice that

$$\widehat{\text{DFPR}}(\rho) = \left[1 - \frac{\widehat{\text{ATS}}_0(\rho) - \widehat{\overline{T}}_{0,l}}{\widehat{\overline{T}}_{0,r} - \widehat{\overline{T}}_{0,l}}\right] \times \widehat{\text{FPR}}(\rho), \tag{26}$$

where $\widehat{\overline{T}}_{0,r} = \max_{1 \leq i \leq n_{\bar{D}}} t_{i,J}$ and $\widehat{\overline{T}}_{0,l} = \min_{1 \leq i \leq n_{\bar{D}}} t_{i,J}$. Similarly,

$$\widehat{\text{DTPR}}(\rho) = \left[1 - \frac{\widehat{\text{ATS}}_1(\rho) - \widehat{\overline{T}}_{1,l}}{\widehat{\overline{T}}_{1,r} - \widehat{\overline{T}}_{1,l}}\right] \times \widehat{\text{TPR}}(\rho), \tag{27}$$

where $\widehat{\overline{T}}_{1,r} = \max_{1 \leq i \leq n_D} t_{i,J}$ and $\widehat{\overline{T}}_{1,l} = \min_{1 \leq i \leq n_D} t_{i,J}$. By combining (26)-(27) with the results in (20) and (21), we have $\widehat{\text{DTPR}}(\rho) \xrightarrow{p} \text{DTPR}(\rho)$ and $\widehat{\text{DFPR}}(\rho) \xrightarrow{p} \text{DFPR}(\rho)$. Because $\widehat{\text{PM-ROC}}(\rho) = (\widehat{\text{DFPR}}(\rho), \widehat{\text{DTPR}}(\rho))$ and $\text{PM-ROC}(\rho) = (\text{DFPR}(\rho), \text{DTPR}(\rho))$, we have $\|\widehat{\text{PM-ROC}}(\rho) - \text{PM-ROC}(\rho)\|^2 = \left(\widehat{\text{DTPR}}(\rho) - \text{DTPR}(\rho)\right)^2 + \left(\widehat{\text{DFPR}}(\rho) - \text{DFPR}(\rho)\right)^2$. Thus, by combining the results in the previous two sentences, we have $\widehat{\text{PM-ROC}}(\rho) \xrightarrow{p} \text{PM-ROC}(\rho)$, for each $\rho \in R^{+}$. Therefore, the result in (22) is valid.

C Proof of Theorem 2

Without loss of generality, we prove the results in cases when $\overline{T}_{0,l} = \overline{T}_{1,l} = 0$. Otherwise, these quantities can be subtracted from the observation times. Let $\Omega_1 = \{\omega_1 | \omega_1 \in R\}$, $\mathcal{B}_1$ be the Borel $\sigma-$algebra on Ω_1, and $\mathbb{P}_1$ be the probability distribution $\text{ATS}(\rho)/T_r$, where $T_r = \lim_{\rho\to\infty} \text{ATS}(\rho)$. Then, $(\Omega_1, \mathcal{B}_1, \mathbb{P}_1)$ is a well defined probability space. Let $\Omega_2 = \{\omega_2 | \omega_2 = \{y(t), t \in [0, T_r]\}$, $\mathcal{B}_2$ be the Borel $\sigma-$algebra on Ω_2, and $\mathbb{P}_2$ be probability measure generated by $\{y(t), t \in [0, T_r]\}$. Then, $(\Omega_2, \mathcal{B}_2, \mathbb{P}_2)$ is another probability space. Now, let us consider the product space $\Omega = \Omega_1 \times \Omega_2$. In this space, define the product Borel $\sigma-$algebra $\mathcal{B} = \mathcal{B}_1 \times \mathcal{B}_2$. The probability measure $\mathbb{P}$ on the space Ω is defined in the way that its marginal probability measures on Ω_1 and Ω_2 are respectively $\mathbb{P}_1$ and $\mathbb{P}_2$, and the conditional probability has the property

$$\widetilde{\mathbb{P}}(A_1|\omega_2) = \frac{T(\rho)}{T_r}$$

where $A_1 = (-\infty, \rho] \in \mathcal{B}_1$ and $\omega_2 \in \Omega_2$. It can be verified that $\widetilde{\mathbb{P}}$ is a transfer probability on $\mathcal{B}_1 \times \Omega_2$. Namely, given $A_1 \in \mathcal{B}_1$, $\widetilde{\mathbb{P}}(A_1|\omega_2) = T(\rho)/T_r$ is a $\mathcal{B}_2-$measurable function, and given $\omega_2 \in \Omega_2$, $\widetilde{\mathbb{P}}(A_1|\omega_2) = T(\rho)/T_r$ is a probability measure on $\mathcal{B}_1$. By the Tuclea Theorem, we can construct a product probability measure based on the transfer probability $\widetilde{\mathbb{P}}$ and the corresponding marginal probability $\mathbb{P}_2$. Now, it can be verified that this product probability has the marginal probability measure $\mathbb{P}_1$ on Ω_1, by noticing

$$\begin{aligned}
\int_{-\infty}^{\infty} \widetilde{\mathbb{P}}(A_1|\omega_2) \mathrm{d}\mathbb{P}_2(\omega_2) &= \int_{-\infty}^{\infty} \frac{T(\rho)}{T_r} \mathrm{d}\mathbb{P}_2(\omega_2) \\
&= \frac{\mathbb{E}(T)}{T_r} \\
&= \frac{\text{ATS}(\rho)}{T_r} = \mathbb{P}_1(A_1).
\end{aligned}$$

Now, let $(\Omega^{(1)}, \mathcal{B}^{(1)}, \mathbb{P}^{(1)})$ and $(\Omega^{(2)}, \mathcal{B}^{(2)}, \mathbb{P}^{(2)})$ be two independent copies of the probability space $(\Omega, \mathcal{B}, \mathbb{P})$ defined above. By the Fubini theorem, we can construct the probability space $(\Omega^*, \mathcal{B}^*, \mathbb{P}^*)$ such that $\Omega^* = \Omega^{(1)} \times \Omega^{(2)}$, $\mathcal{B}^* = \mathcal{B}^{(1)} \times \mathcal{B}^{(2)}$, and

$$\mathbb{P}^*(B_1 \times B_2) = \mathbb{P}^{(1)}(B_1) \times \mathbb{P}^{(2)}(B_2), \quad \text{for any} \quad B_1 \in \mathcal{B}^{(1)} \text{ and } B_2 \in \mathcal{B}^{(2)}.$$

Now, for any $\omega^* = (\omega^{(1)}, \omega^{(2)} = (\omega_1^{(1)}, \omega_2^{(1)}, \omega_1^{(2)}, \omega_2^{(2)})$ in Ω^*, where $\omega^{(1)} = (\omega_1^{(1)}, \omega_2^{(1)}) \in \Omega^{(1)}$ and $\omega^{(2)} = (\omega_1^{(2)}, \omega_2^{(2)}) \in \Omega^{(2)}$, we define a random variable $Y_2^*(\omega^*) = \omega_1^{(2)}$ in the probability space $(\Omega^*, \mathcal{B}^*, \mathbb{P}^*)$. On the other hand, the random variable Y_1^* defined in (5) can be regarded

as a random variable in the same probability space, satisfying $Y_1^*(\omega^*) = g^*(\omega_2^{(1)})$ where g^* is a deterministic function. Because Y_1^* takes the values in $\Omega^{(1)}$ only, and Y_2^* takes the values in $\Omega^{(2)}$ only, they are independent of each other.

Next, we'll verify that Y_2^* defined above satisfy Equations (14) and (15). To this end, we first notice that $\mathbb{P}^{(2)}(Y_2^* \leq \rho|\omega_2^{(2)}) = T(\rho)/T_r$. By the total expectation formula, we have

$$\begin{aligned}\mathbb{P}^*(Y_2^* \leq \rho|D=0) &= \mathbb{E}(\mathbb{P}^{(2)}(A|\omega_2^{(2)})|D=0) \\ &= \mathbb{E}\left(\frac{T(\rho)}{T_r}|D=0\right) \\ &= \frac{\mathrm{ATS}_0(\rho)}{\overline{T}_{0,r}}.\end{aligned}$$

Similarly, we can check that

$$\mathbb{P}^*(Y_2^* \leq \rho|D=1) = \frac{\mathrm{ATS}_1(\rho)}{\overline{T}_{1,r}}.$$

Therefore, Equations (14) and (15) are valid. This concludes the proof.

D Proof of Theorem 3

We have

$$\begin{aligned}\mathrm{DAUC} &= \int_0^1 \mathrm{DTPR}\left(\mathrm{DFPR}^{-1}(t)\right)\mathrm{d}t \\ &= \int_{-\infty}^{\infty} \mathrm{DTPR}(u)\mathrm{dDFPR}(y) \\ &= \int_{-\infty}^{\infty} (1-\mathbb{P}(Y^* \leq u|D=1))\,\mathrm{d}\mathbb{P}(Y^* > u|D=0) \\ &= \mathbb{P}(Y_{\overline{D}}^* \leq Y_D^*).\end{aligned}$$

So, Equation (23) is valid.

References

Chen, K., and Jin, Z. (2005), "Local polynomial regression analysis of clustered data," *Biometrika*, **92**, 59–74.

Cupples, L.A. et al. (2007), "The Framingham Heart Study 100K SNP genome-wide association study resource: overview of 17 phenotype working group reports," *BMC Medical Genetics*, **8**, (Suppl 1): S1, doi://10.1186/1471-2350-8-S1-S1.

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

Heagerty, P.J., Lumley, T., and Pepe, M.S. (2000), "Time-dependent ROC curves for censored survival data and a diagnostic marker," *Biometrics*, **56**, 337–344.

Heagerty, P.J., and Zheng, Y. (2005), "Survival model predictive accuracy and ROC curves," *Biometrics*, **61**, 92–105.

Li, Y. (2011), "Efficient semiparametric regression for longitudinal data with nonparametric covariance estimation," *Biometrika*, **98**, 355–370.

Li, J., and Qiu, P. (2016), "Nonparametric dynamic screening system for monitoring correlated longitudinal data," *IIE Transactions*, **48**, 772–786.

Li, J., and Qiu, P. (2017), "Construction of an efficient multivariate dynamic screening system," *Quality and Reliability Engineering International*, **33**, 1969–1981.

Li, S.Y., Tang, L.C., and Ng, S.H. (2010), "Nonparametric CUSUM and EWMA control charts for detecting mean shifts," *Journal of Quality Technology*, **42**, 209–226.

Liang, K.Y., and Zeger, S.L. (1986), "Longitudinal data analysis using generalized linear models," *Biometrika*, **73**, 13–22.

Lin, X., and Carroll, R. (2000), "Nonparametric function estimation for clustered data when the predictor is measured without/with error," *Journal of the American Statistical Association*, **95**, 520–534.

Ma, S., Yang, L., and Carroll, R. (2012), "A simultaneous confidence band for sparse longitudinal regression," *Statistica Sinica*, **22**, 95–122.

Obuchowski, N.A. (2003), "Receiver operating characteristic curves and their use in radiology," *Radiology,* **229**, 3–8.

Pepe, M.S. (2003), *The Statistical Evaluation of Medical Tests for Classification and Prediction,* New York: Oxford University Press.

Qiu, P. (2014), *Introduction to Statistical Process Control*, Boca Raton, FL: Chapman & Hall/CRC.

Qiu, P. (2018), "Some perspectives on nonparametric statistical process control," *Journal of Quality Technology*, **50**, 49–65.

Qiu, P., and Le, Chap (2001), "ROC curve estimation based on local smoothing," *Journal of Statistical Computation and Simulation*, **70**, 55–69.

Qiu, P., and Xiang, D. (2014), "Univariate dynamic screening system: an approach for identifying individuals with irregular longitudinal behavior," *Technometrics*, **56**, 248–260.

Qiu, P., and Xiang, D. (2015), "Surveillance of cardiovascular diseases using a multivariate dynamic screening system," *Statistics in Medicine*, **34**, 2204–2221.

Qiu, P., Zi, X., and Zou, C. (2018), "Nonparametric dynamic curve monitoring," *Technometrics*, **60**, 386–397.

Saha, P., and Heagerty, P.J. (2010), "Time-dependent predictive accuracy in the presence of competing risks," *Biometrics*, **66**, 999–1011.

Wang, N. (2003), "Marginal nonparametric kernel regression accounting for within-subject correlation," *Biometrika*, **90**, 43–52.

Xiang, D., Qiu, P., and Pu, X. (2013), "Local polynomial regression analysis of multivariate longitudinal data," *Statistica Sinica*, **23**, 769–789.

You, L., and Qiu, P. (2019), "Fast computing for dynamic screening systems when analyzing correlated data," *Journal of Statistical Computation and Simulation*, **89**, 379–394.

Zhao, Z., and Wu, W. (2008), "Confidence bands in nonparametric time series regression," *Annals of Statistics*, **36**, 1854–1878.

Zhou, X.H., Obuchowski, N.A., and McClish, D.K. (2002), *Statistical Methods in Diagnostic Medicine (1st ed.)*, New York: John Wiley & Sons.