

Early Detection of High Disease Activity in Juvenile Idiopathic Arthritis by Sequential Monitoring of Patients' Health-Related Quality of Life Scores

Lu You¹, Andrew Qiu², Bin Huang³, and Peihua Qiu¹

¹Department of Biostatistics, University of Florida, Gainesville, FL 32610

²Buchholz High School, Gainesville, FL 32606

³Department of Environmental Health, University of Cincinnati, Cincinnati, OH 45229

Abstract

Juvenile idiopathic arthritis (JIA) is a chronic disease. During its “high disease activity (HDA)” stage, JIA can cause severe pain, and thus could seriously affect patients’ physical and psychological health. Early detection of the HDA stage of JIA can reduce the damage of the disease by treating it at an early stage and alleviating the painful experience of the patients. So far, no effective cure of JIA has been found, and one major goal of disease management is to improve patients’ quality of life. To this end, patients’ health-related quality of life scores are routinely collected over time from JIA patients. In this paper, we demonstrate that a new statistical methodology called dynamic screening system (DySS) is effective for early detection of the HDA stage of JIA. By this approach, a patient’s health-related quality of life scores are monitored sequentially, and a signal is given by DySS once the longitudinal pattern of the scores is found to be significantly different from the pattern of patients with low disease activity. Dimension reduction of the observed health-related quality of life scores and the corresponding impact on the performance of DySS are also discussed.

Key Words: Juvenile idiopathic arthritis; Longitudinal data; Online monitoring; Quality of life; Statistical process control; Model selection.

1 Introduction

Juvenile idiopathic arthritis (JIA) is a common chronic rheumatic disease among children and adolescents (cf., Giancane et al. 2016). It can cause physical pain, disability, and even morbidity,

and thus affects patients' daily life and well-being in a significant way. Currently, the etiology of JIA is still unknown, and no effective treatments have been found. Therefore, a main objective of JIA disease management and treatment is to alleviate symptoms and improve patients' overall quality of life (cf., Consolaro et al. 2016).

In the past few years, numerous outcome measures have been proposed to assess JIA children's condition. First proposed in Consolaro et al. (2009), the juvenile arthritis disease activity score (JADAS) is a composite score based on four sub-measures. In the literature, the validity of JADAS has been clinically investigated by different researchers (cf., Mcerlane et al. 2013), and JADAS was shown to be more effective than measures based solely on counting disease-active joints (cf., Wu et al. 2016). However, to obtain a JADAS score, patients need to visit clinics for their disease status to be evaluated by physicians. Furthermore, calculation of JADAS also involves a blood test, which would usually take a day or two for results. Thus, it is often inconvenient to obtain the JADAS score in a routine basis. On the other hand, health-related quality of life (HRQOL) scores have been found to be the most relevant patient-reported outcomes (PROs) for assessment of patients' suffering from chronic diseases, including JIA (cf., Mańczak et al. 2016). HRQOL is a comprehensive and multi-dimensional concept that encompasses physical, mental, emotional, and social elements associated with health (cf., WHOQOL Group, 1998). Among several other instruments for evaluating HRQOL, pediatric quality of life inventory (PedsQL) is a commonly used questionnaire-based model for evaluating quality of life of patients aging between 2 to 18 (cf., Varni et al. 2002a). PedsQL generic core scales (version 4.0) measure the overall quality-of-life from four different dimensions suggested by the World Health Organization. PedsQL rheumatology module (version 3.0) is another approach designed specifically to study the quality of life of JIA children (cf., Varni et al. 2002b). Compared to clinical assessment by physicians, HRQOL can measure certain factors that may be overlooked by physicians, such as medication side effects, fear of injections, and missed school (cf., Weiss et al. 2013), and is easier to use for self-reporting and self-monitoring purposes.

Among all JIA patients, those at the "high disease activity (HDA)" stage suffer significantly

more from the disease. Thus, it is important to distinguish them from other patients early on. To this end, a traditional approach is based on the JADAS scores. As discussed above, this approach is often time consuming, costly, and inconvenient to patients, because it is based on clinic visits and blood tests. In comparison, HRQOL scores are easier and more economic to collect. Therefore, a natural research question is whether JIA patients at the HDA stage can be detected early based on the observed HRQOL scores. Recent research in a new statistical area, called dynamic screening system (DySS), makes it possible for us to answer this question (Qiu and Xiang 2014, 2015, Li and Qiu 2016, 2017, Qiu et al. 2018). By a DySS approach, the longitudinal pattern of the HRQOL scores for JIA patients at the “low disease activity (LDA)” stage can be first estimated properly from a dataset containing observed HRQOL scores of JIA patients whose LDA status has been confirmed by the traditional JADAS scores. This dataset is called an in-control (IC) dataset in this paper and the estimated longitudinal pattern is called the regular longitudinal pattern, to make the terminologies consistent with those in the statistical process control (SPC) literature (cf., Qiu 2014). Then, the observed HRQOL scores of a new JIA patient are monitored sequentially, and their longitudinal pattern is compared to the regular longitudinal pattern over time. A signal will be given once the cumulative difference between the two longitudinal patterns exceeds a threshold value, and the specific JIA patient is declared to be at the HDA status after receiving the signal. In this paper, we propose several modifications of some existing DySS methods for the JIA application, and compare their performance using a real JIA data and some simulation studies.

The remaining parts of the article are organized as follows. Some existing DySS methods and their modifications are first introduced briefly in Section 2. Then, a JIA dataset is analyzed by these DySS methods in Section 3. Some simulation results are provided in Section 4 to further evaluate their numerical performance. Finally, some remarks conclude the article in Section 5.

2 A Brief Description of the DySS Approach

The DySS approach consists of two major steps. First, the regular longitudinal pattern of the predictors (i.e., the HRQOL scores in the JIA problem) is estimated using longitudinal data modeling.

Then, for a specific patient, the standardized difference between his/her longitudinal pattern of the predictors and the estimated regular pattern is monitored sequentially by a control chart and a signal will be given by the chart once the cumulative difference exceeds a threshold. These two steps are described briefly below in two subsections.

2.1 Estimation of the regular longitudinal pattern of the predictors

Assume that the IC dataset contains observed longitudinal data of the q -dimensional vector of predictors, \mathbf{y} , and the observed data follow the model

$$\mathbf{y}_i(t_{ij}) = \boldsymbol{\mu}(t_{ij}) + \boldsymbol{\epsilon}_i(t_{ij}), \quad \text{for } j = 1, 2, \dots, m_i, \quad i = 1, 2, \dots, n, \quad (1)$$

where $t_{ij} \in [0, T]$ is the j th observation time of the i th patient, $\boldsymbol{\mu}(t_{ij}) = (\mu_1(t_{ij}), \dots, \mu_q(t_{ij}))'$ is the mean of $\mathbf{y}_i(t_{ij}) = (y_{i,1}(t_{ij}), \dots, y_{i,q}(t_{ij}))'$, and $\boldsymbol{\epsilon}_i(t_{ij})$ is a vector of random noise. The data observed from different patients are assumed to be independent, and the ones from a same patient are assumed to be correlated with the correlation matrix function being $\text{Cov}(\boldsymbol{\epsilon}_i(s), \boldsymbol{\epsilon}_i(t)) = \boldsymbol{\Sigma}(s, t) = (\sigma_{k_1 k_2}(s, t))_{k_1, k_2=1, \dots, q}$, for $s, t \in [0, T]$. The regular longitudinal pattern of $\mathbf{y}_i(t)$ is then described by $\boldsymbol{\mu}(t)$ and $\boldsymbol{\Sigma}(s, t)$.

By the nonparametric estimation approach suggested in Xiang et al. (2013) for multivariate longitudinal data modeling, $\boldsymbol{\mu}(t)$ can be estimated by

$$\hat{\boldsymbol{\mu}}(t) = \hat{\boldsymbol{\beta}}'(\mathbf{I}_{q \times q} \otimes \mathbf{e}_1), \quad (2)$$

where

$$\hat{\boldsymbol{\beta}} = \left[\sum_{i=1}^n (\mathbf{I}_{q \times q} \otimes \mathbf{X}_i)' \mathbf{W}_i (\mathbf{I}_{q \times q} \otimes \mathbf{X}_i) \right]^{-1} \left[\sum_{i=1}^n (\mathbf{I}_{q \times q} \otimes \mathbf{X}_i)' \mathbf{W}_i \mathbf{Y}_i \right],$$

$\mathbf{I}_{q \times q}$ is the $q \times q$ identity matrix, $\mathbf{e}_1 = (1, 0, \dots, 0)'$ is the q -dimensional unit vector, $\mathbf{W}_i = \text{diag}\{\mathbf{K}_{ij}, j = 1, \dots, m_i\}$, $\mathbf{K}_{ij} = \text{diag}\{K((t_{ij} - t)/h_l)/h_l, l = 1, \dots, q\}$, $\{h_l\}$ are properly chosen bandwidths, $K(u) = 3(1 - u^2)I(|u| \leq 1)/4$ is the Epanechnikov kernel function, \mathbf{X}_i are the $m_i \times 2$ design matrices with the first column being $(1, \dots, 1)'$ and the second column being $((t_{i1} - t), \dots, (t_{im_i} - t))'$, and $\mathbf{Y}_i = (y_{i,1}(t_{i1}), \dots, y_{i,1}(t_{im_i}), \dots, y_{i,q}(t_{i1}), \dots, y_{i,q}(t_{im_i}))'$. Let

$\widehat{\boldsymbol{\epsilon}}_i(t_{ij}) = \mathbf{y}_i(t_{ij}) - \widehat{\boldsymbol{\mu}}(t_{ij}) = (\widehat{\epsilon}_{ij,1}, \dots, \widehat{\epsilon}_{ij,q})'$ be the residuals. Then, the (k_1, k_2) -th element $\sigma_{k_1, k_2}(s, t)$ of $\boldsymbol{\Sigma}(s, t)$ can be estimated by the following kernel estimator

$$\widehat{\sigma}_{k_1, k_2}(s, t) = \frac{\sum_{i=1}^n \sum_{j_1=1}^{m_i} \sum_{j_2=1}^{m_i} \widehat{\epsilon}_{ij_1, k_1} \widehat{\epsilon}_{ij_2, k_2} K((t_{ij_1} - s)/g_{l_1}) K((t_{ij_2} - t)/g_{l_2})}{\sum_{i=1}^n \sum_{j_1=1}^{m_i} \sum_{j_2=1}^{m_i} K((t_{ij_1} - s)/g_{l_1}) K((t_{ij_2} - t)/g_{l_2})}, \quad (3)$$

where $\{g_l, l = 1, \dots, q\}$ are bandwidths that could be chosen different from $\{h_l, l = 1, \dots, q\}$. Selection of the bandwidths has been discussed in detail in Qiu and Xiang (2015).

2.2 Online dynamic monitoring

There are several different versions of DySS for online monitoring of a new patient. Some major ones are described below. The DySS procedure discussed in Qiu and Xiang (2014) is for univariate cases when $q = 1$. In such cases, there is a single predictor $y(t)$. Assume that $y(t)$ is observed at times $t_1^*, t_2^*, \dots \in [0, T]$ for a new patient under monitoring, where the observation times $\{t_1^*, t_2^*, \dots\}$ could be different from those in the IC data (cf., model (1)). Then, the standardized observations are defined by

$$e_j = \frac{y(t_j^*) - \widehat{\mu}(t_j^*)}{\widehat{\sigma}(t_j^*)}, \quad \text{for } j = 1, 2, \dots, \quad (4)$$

where $\widehat{\mu}(t_j^*)$ and $\widehat{\sigma}(t_j^*) = \sqrt{\widehat{\Sigma}(t_j^*, t_j^*)}$ are obtained from (2) and (3). To detect a downward mean shift in $y(t)$, we can use the following downward CUSUM charting statistic

$$C_j^- = \min\{0, C_{j-1}^- + e_j + k\}, \quad \text{for } j \geq 1, \quad (5)$$

where $C_0^- = 0$, and $k > 0$ is an allowance constant. The chart gives a signal when $C_j^- < \rho_C$, for some control limit $\rho_C < 0$. The corresponding upward and two-sided CUSUM charts can be defined similarly. Because we are mainly interested in detecting low HRQOL scores in the JIA data, the downward CUSUM chart is focused in this paper. From (4) and (5), it can be seen that the CUSUM chart (5) uses the cumulative difference between the estimated regular longitudinal pattern of $y(t)$ described by $\widehat{\mu}(t_j^*)$ and $\widehat{\sigma}(t_j^*)$ and the observed longitudinal pattern of the new patient, and gives a signal when the cumulative difference exceeds the control limit ρ_C downward. From (5), it can also be seen that the CUSUM chart has the re-starting mechanism that its charting statistic is reset to 0 each time when the cumulative difference by the current time t_j^* suggests no downward shift in

the mean of $y(t)$, implied by the fact that $C_{j-1}^- + e_j \geq -k$. This re-starting mechanism makes the CUSUM chart possess some good theoretical properties (Moustakides 1986).

Instead of the CUSUM chart (5), some researchers prefer to use the exponentially weighted moving average (EWMA) chart defined by

$$E_j = \lambda e_j + (1 - \lambda)E_{j-1}, \quad \text{for } j \geq 1, \quad (6)$$

where $E_0 = 0$, and $\lambda \in (0, 1]$ is a weighting parameter. The chart gives a signal of downward mean shift in $y(t)$ at the time t_j^* if $E_j < \rho_E$, where $\rho_E < 0$ is a control limit. Mathematically, the EWMA chart (6) is easy to study because E_j is a weighted average of $\{e_1, e_2, \dots, e_j\}$. In the literature, it has been well demonstrated that it has a comparable performance to the CUSUM chart (5) if its weighting parameter λ is chosen properly (cf., Section 5.2.3, Qiu 2014).

The charts (5) and (6) are developed mainly for cases when the observation distribution is normal. When the observation distribution is non-normal, there are a number of nonparametric control charts available in the literature. See a recent overview on this topic in Qiu (2018). Here, we adapt the Wilcoxon rank-sum based nonparametric EWMA chart by Li et al. (2010) to the DySS problem, which is described below. Let $R = \{e_{ij} : e_{ij} = [y_i(t_{ij}) - \hat{\mu}(t_{ij})]/\hat{\sigma}(t_{ij}), j = 1, \dots, m_i, i = 1, \dots, n\}$ be the set of standardized observations of the patients in the IC data, and $n_R = \sum_{i=1}^n m_i$ be its cardinality. To detect a downward shift in $y(t)$, we first calculate the rank of the standardized observation e_j of the patient under monitoring in the dataset R , by $W_j = \sum_{r \in R} I(e_j \geq r) + 1$. The quantity W_j is also called the Wilcoxon rank-sum statistic of e_j in comparison with R . Then, the following EWMA charting statistic is considered:

$$\tilde{E}_j = \lambda(W_j - \tilde{\mu}_{W_j})/\tilde{\sigma}_{W_j} + (1 - \lambda)\tilde{E}_{j-1} \quad \text{for } j \geq 1, \quad (7)$$

where $\tilde{E}_0 = 0$, $\lambda \in (0, 1]$ is a weighting parameter, $\tilde{\mu}_{W_j} = (n_R + 1)/2$ is the IC mean of W_j , and $\tilde{\sigma}_{W_j}^2 = n_R(n_R + 2)/12$ is its IC variance. The chart gives a signal at time t_j^* if $\tilde{E}_j < \rho_{\tilde{E}}$, where $\rho_{\tilde{E}}$ is a control limit.

In multivariate cases when $q > 1$, Qiu and Xiang (2015) suggested using the following multi-

variate EWMA chart:

$$\mathbf{E}_j = \lambda \mathbf{e}_j + (1 - \lambda) \mathbf{E}_{j-1}, \quad \text{for } j \geq 1, \quad (8)$$

where $\mathbf{E}_0 = \mathbf{0}$, \mathbf{e}_j is the standardized observation defined by

$$\mathbf{e}_j = \widehat{\boldsymbol{\Sigma}}(t_j^*, t_j^*)^{-1/2} [\mathbf{y}(t_j^*) - \widehat{\boldsymbol{\mu}}(t_j^*)], \quad \text{for } j = 1, 2, \dots,$$

and $\lambda \in (0, 1]$ is a weighting parameter. The chart gives a signal at t_j^* if

$$\frac{2 - \lambda}{\lambda} \mathbf{E}_j' \mathbf{E}_j > \rho_M, \quad (9)$$

where $\rho_M > 0$ is a control limit. Like many other multivariate SPC charts, the chart (8)-(9) is designed for detecting mean shifts in all directions. In the JIA example discussed in Section 1, however, we are mainly concerned about cases when one or more components of $\mathbf{y}(t)$ have lower mean values, compared to the estimated regular longitudinal pattern obtained from the IC data. For this reason, we can consider an alternative chart for detecting one-sided shifts. Let $\mathbf{e}_j = (e_{j,1}, \dots, e_{j,q})'$, and

$$E_{j,k} = \lambda e_{j,k} + (1 - \lambda) E_{j-1,k}, \quad \text{for } j \geq 1, \text{ and } k = 1, \dots, q, \quad (10)$$

where $E_{0,k} = 0$. Then, the chart gives a signal at t_j^* if $\min_{k=1, \dots, q} E_{j,k} < \rho_O$, where $\rho_O < 0$ is a pre-specified control limit. This chart (10) is different from the one by Woodall and Ncube (1985) in that the former accounts for the between variable correlation by transforming the original observations $\mathbf{y}(t)$ into the standardized observations $\mathbf{e}(t_j^*) = \widehat{\boldsymbol{\Sigma}}(t_j^*, t_j^*)^{-1/2} (\mathbf{y}(t_j^*) - \widehat{\boldsymbol{\mu}}(t_j^*))$, which are uncorrelated among the q components, while the latter does not.

To accommodate serial data correlation, Li and Qiu (2016,2017) proposed to decorrelate serial observations by Cholesky decomposition before monitoring the standardized observations. The one by Li and Qiu (2016) is on univariate cases, which is briefly described below. For multivariate cases, see Li and Qiu (2017). In univariate cases when $q = 1$, let $\boldsymbol{\epsilon}_j = (\epsilon(t_1^*), \dots, \epsilon(t_j^*))'$ be the vector of all random errors in model (1) by the current time point t_j^* . Its estimated variance-covariance matrix is denoted as $\widehat{\boldsymbol{\Sigma}}_{jj}$, where

$$\widehat{\boldsymbol{\Sigma}}_{jj} = \begin{pmatrix} \widehat{\boldsymbol{\Sigma}}_{j-1,j-1} & \widehat{\boldsymbol{\sigma}}_{j-1,j} \\ \widehat{\boldsymbol{\sigma}}'_{j-1,j} & \widehat{\sigma}_{jj} \end{pmatrix}, \quad \text{for } j \geq 2.$$

Then, the standardized observations e_1, \dots, e_j are defined by

$$\begin{cases} e_1 = \widehat{\epsilon}(t_1^*)/d_1 \\ e_j = \frac{-\widehat{\boldsymbol{\sigma}}'_{j-1,j} \widehat{\boldsymbol{\Sigma}}_{j-1,j-1}^{-1} \widehat{\boldsymbol{\epsilon}}_{j-1} + \widehat{\epsilon}(t_j^*)}{d_j}, \quad \text{for } j \geq 2, \end{cases}$$

where $d_1 = \sqrt{\widehat{\Sigma}_{11}}$, $d_j = \sqrt{\widehat{\sigma}_{jj} - \widehat{\boldsymbol{\sigma}}'_{j-1,j} \widehat{\boldsymbol{\Sigma}}_{j-1,j-1}^{-1} \widehat{\boldsymbol{\sigma}}_{j-1,j}}$, for $j \geq 2$, and $\widehat{\boldsymbol{\epsilon}}_j = (\widehat{\epsilon}(t_1^*), \dots, \widehat{\epsilon}(t_j^*))'$ are the residuals of the estimated model of (1). It has been confirmed in Li and Qiu (2016) that e_1, \dots, e_j are asymptotically uncorrelated with a common mean of 0 and a common variance of 1. The CUSUM chart (5) can then be applied to the decorrelated data e_1, e_2, \dots . Similarly, in multivariate cases when $q \geq 2$, the EWMA charts (8)-(9) and (10) can be applied to the decorrelated observations.

In the literature, Fricker (2007) suggested a one-sided multivariate CUSUM chart for disease screening. That CUSUM chart was modified from the one by Crosier (1986), and it was designed for cases when the IC mean was a constant over time. Next, we modify that one-sided CUSUM chart in the framework of DySS for detecting downward shifts, and the modified chart will be considered in the analysis of JIA data discussed in the next section. Let $\mathbf{S}_0 = \mathbf{0}$, and for $j \geq 1$,

$$\begin{cases} \mathbf{S}_j = \mathbf{0}, & \text{if } C_j \leq k, \\ \mathbf{S}_j = \min \{(\mathbf{S}_{j-1} + \mathbf{y}(t_j^*) - \widehat{\boldsymbol{\mu}}(t_j^*))(1 - k/C_j), \mathbf{0}\}, & \text{otherwise,} \end{cases} \quad (11)$$

where \min is an element-wise operator (i.e., taking the minimum of the corresponding elements of the vectors), $C_j = [(\mathbf{S}_{j-1} + \mathbf{y}(t_j^*) - \widehat{\boldsymbol{\mu}}(t_j^*))' \widehat{\boldsymbol{\Sigma}}(t_j^*, t_j^*)^{-1} (\mathbf{S}_{j-1} + \mathbf{y}(t_j^*) - \widehat{\boldsymbol{\mu}}(t_j^*))]^{1/2}$, k is an allowance constant, and $\widehat{\boldsymbol{\mu}}(t)$ and $\widehat{\boldsymbol{\Sigma}}(s, t)$ are the estimated IC mean vector and IC correlation matrix function obtained from (2) and (3). The chart gives a signal if $[\mathbf{S}'_j \widehat{\boldsymbol{\Sigma}}(t_j^*, t_j^*)^{-1} \mathbf{S}_j]^{1/2} > \rho_F$ for a pre-specified control limit ρ_F . Unlike the original one-sided multivariate CUSUM chart by Fricker (2007), this chart (11) has taken into account the dynamic nature of the IC process distribution and the between-variable correlation. At the same time, it keeps the one-sided nature of the chart in Fricker (2007) that it is sensitive to the downward shift in each component of $\mathbf{y}(t_j^*)$ only.

3 Application to a JIA data

3.1 The JIA data

The JIA data were obtained from a prospective cohort study of the Cincinnati Children’s Hospital that included 530 young JIA patients of 2-16 years old. Those who would suffer severely from the medical diagnoses of JIA disease conditions have been excluded from the study. Patients’ disease conditions were evaluated using the JADAS approach that included the following four measures: (i) physician’s global assessment of disease activity, (ii) parent global assessment of well-being, (iii) erythrocyte sedimentation rate (ESR), and (iv) a count of joints with active disease up to a maximum of ten joints. The version of JADAS used in the study is usually referred to as JADAS10 in the literature. All four measures in JADAS were scaled to the range of 0 to 10, with 0 being the best and 10 being the worst. The JADAS score is defined to be the summation of the four measures; thus, it ranges between 0 and 40.

Patients’ health-related quality of life (HRQOL) was evaluated over time using the PedsQL generic core scales (version 4.0) and the PedsQL rheumatology module (version 3.0). The PedsQL generic core scales are designed for measuring the core health dimensions outlined by the World Health Organization, including the following four dimensions: physical functioning, emotional functioning, social functioning, and school functioning (cf., Varni et al. 2002a). A total score is calculated by averaging the four sub-scores. The PedsQL rheumatology module is a disease-specific module that considers five disease-related dimensions, including pain and hurt, daily activities, treatment, worry, and communication (cf., Varni et al. 2002a). A rheumatology total score is also calculated by averaging the five sub-scores. All PedsQL scores range from 0 to 100, with higher scores indicating better HRQOL.

Varni et al. (2002b) and Weiss et al. (2013) have observed association between PedsQL generic core scales and rheumatology module. Varni et al. (2002b) also explored the possibility to use the rheumatology module for detecting clinical condition changes over time. They believed that the rheumatology module should be more sensitive to patients’ health conditions since it was developed

for measuring the HRQOL dimensions that were specifically related to JIA. Seid et al. (2009) also found some correlation between JIA disease activity and HRQOL, although they could not explain a considerable remaining variation in HRQOL. Other related research in the literature includes those in Weiss et al. (2013) and Seid et al. (2014) for exploring the possible relationship between HRQOL and the JIA medical condition and certain nonmedical factors, such as self-efficacy, coping with pain, barriers to adherence, social support, and parental distress. *However, all these existing research is retrospective, and there does not exist a single method for prospective online monitoring of JIA patients' medical condition from their HRQOL scores yet. We will show in the next part that the recent statistical methodology DySS can serve that purpose well.*

3.2 Detection of JIA high disease activity by DySS

In the JIA data, all 530 patients have been diagnosed to have JIA. Some patients' JIA conditions were more severe than those of the others. As discussed in Section 1, there are no existing medical treatments that can cure JIA completely, and the current disease management and treatment is mainly for alleviating the disease symptoms and improving patients' quality of life. To this end, it is important to know whether the disease condition of a JIA patient is in high disease activity (HDA) stage or in low disease activity (LDA) stage, because the corresponding strategy for disease management and treatment would be different for different disease conditions. Based on data of "more than 9,000 patients with JIA from 118 pediatric rheumatology centres in 49 countries," Consolaro et al. (2017) suggested using the 75th percentile of the JADAS scores as a cut point for defining the HDA and LDA disease conditions for two major JIA disease types oligoarthritis and RF-negative polyarthritis. Namely, JIA patients with the top 25 percent of JADAS scores are classified as being in HDA disease condition and the remaining 75 percent are classified as being in LDA disease condition. This classification method has been commonly used in the literature. By this method, the cut points of the JADAS scores for these two disease types are found to be 16.4 and 18.9, respectively. In this paper, we use the same approach to find the cut point of the combination of all remaining JIA disease types, and the cut point is found to be 19.0. The remaining JIA

disease types are considered as a collection here because each individual type is quite rare. A JIA patient is classified into the LDA group if s/he is never in the HDA status in the entire follow-up period. Of all 530 JIA patients, 62 of them have missing JADAS scores. Thus, they are excluded from our data analysis. Among the remaining 468 patients, by using the classification approach described above, 342 are classified into the LDA group and 126 are classified into the HDA group. To apply the DySS approaches discussed in Section 2, we need an IC dataset for estimating the regular longitudinal pattern. To this end, we randomly select half JIA patients (i.e., 171 patients) from all 342 patients in the LDA group, and the data of the selected 171 patients are used as the IC data. The data of the remaining 171 patients in the LDA group and the 126 patients in the HDA group are used as test data. In this section, patients in the LDA group are called IC patients, and those in the HDA group are called out-of-control (OC) patients, to make the terminologies consistent with those in the SPC literature (cf., Qiu 2014). Figure 1 shows the observed HRQOL scores for both the IC and OC patients, along with their means \pm standard errors (the blue and red shaded regions in the plots). From the plots, it can be seen that the mean HRQOL scores of the IC patients are higher than those of the OC patients, implying that the observed HRQOL scores indeed carry useful information about the disease condition of the JIA patients.

Now, we apply the DySS procedures discussed in Section 2 to the JIA data. For a given DySS procedure with a fixed control limit, its performance can be evaluated by the true positive rate (TPR) and the false positive rate (FPR), where TPR is defined to be the proportion of OC patients who receive signals from the related control chart, and FPR is defined to be the proportion of IC patients who receive signals from the control chart. More specifically, let $D = 0$ denote an IC patient, $D = 1$ denote an OC patient, $S = 0$ denote the patient who does not receive any signal from the control chart, and $S = 1$ denote the patient who receives a signal. Then, for a given control limit ρ , $\text{TPR}(\rho) = P(S = 1|D = 1)$ and $\text{FPR}(\rho) = P(S = 1|D = 0)$. When ρ changes, the curve of the pair $(\text{FPR}(\rho), \text{TPR}(\rho))$ is called the receiver operating characteristic (ROC) curve in the literature (e.g., Pepe 2010). The area under the curve (AUC) refers to the area under the ROC curve that can be computed by $\text{AUC} = \int_0^1 \text{TPR}(\text{FPR}^{-1}(x)) dx$. In medical diagnosis, ROC curve is a popular tool for evaluating the performance of a diagnostic (or classification) method, or for

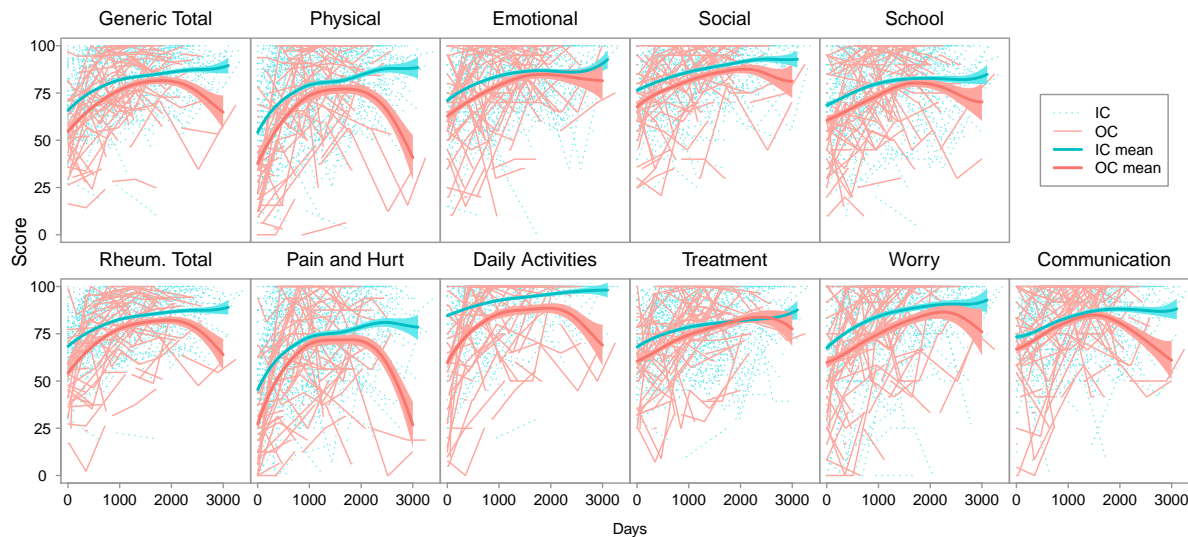


Figure 1: Each plot shows the HRQOL scores of the IC and OC patients observed over time, where blue dashed lines are for IC patient and red solid lines are for OC patients. The IC and OC means and their standard errors are denoted by the blue and red shaded regions. Plots in the first row are for the generic measures and their total (1st column), and those in the second row are for the rheumatology measures and their total (1st column).

comparing two diagnostic methods. By this tool, the DySS method is better than random guessing if its AUC is larger than 0.5, and it performs better if its AUC value is larger.

First, we consider the univariate CUSUM chart (5), the univariate EWMA chart (6), their corresponding versions after serial data decorrelation, and the nonparametric EWMA chart (7). These five charts are denoted as QX-CUSUM, QX-EWMA, LQ-CUSUM, LQ-EWMA, and NP-EWMA, respectively. These charts are used for monitoring the JIA disease status from either the generic total score or the rheumatology total score. In the two CUSUM charts, the allowance constant is chosen to be 0.1, 0.2 and 0.5. The weighting parameter λ is also chosen to be 0.1, 0.2 and 0.5 in the three EWMA charts. For each chart, its AUC value is estimated using bootstrap, described briefly below. Assume that there are n_0 IC patients and n_1 OC patients in the test data. Then, we draw with replacement n_0 patients from the IC patients and n_1 patients from the OC patients. The patients chosen constitute a bootstrap sample. Then, an estimate of AUC can be

Table 1: Estimates of AUC values and the corresponding standard errors (in parentheses) of the five charts QX-CUSUM, QX-EWMA, LQ-CUSUM, LQ-EWMA, and NP-EWMA when monitoring the generic total scores and the rheumatology total scores of the JIA patients in the test data.

		QX-EWMA	QX-CUSUM	LQ-EWMA	LQ-CUSUM	NP-EWMA
$\lambda = k = 0.1$	Generic	0.6114 (0.0013)	0.6289 (0.0012)	0.5811 (0.0015)	0.5776 (0.0012)	0.6073 (0.0013)
	Rheumatology	0.6381 (0.0012)	0.6401 (0.0011)	0.6102 (0.0015)	0.5883 (0.0011)	0.6379 (0.0012)
$\lambda = k = 0.2$	Generic	0.6122 (0.0013)	0.6257 (0.0012)	0.5823 (0.0015)	0.5791 (0.0012)	0.6087 (0.0013)
	Rheumatology	0.6401 (0.0012)	0.6383 (0.0011)	0.6084 (0.0015)	0.5861 (0.0011)	0.6388 (0.0012)
$\lambda = k = 0.5$	Generic	0.6170 (0.0013)	0.6226 (0.0012)	0.5872 (0.0015)	0.5837 (0.0011)	0.6145 (0.0013)
	Rheumatology	0.6446 (0.0012)	0.6405 (0.0011)	0.6133 (0.0015)	0.6004 (0.0011)	0.6434 (0.0012)

computed from the bootstrap sample after the corresponding ROC curve is computed, as described above. This process is repeated 1000 times, and 1000 estimated values of AUC can be obtained accordingly. The average of the 1000 estimated values of AUC is used as the final estimate of AUC, and the standard error of the final estimate can be computed as well. The results of the five charts are presented in Table 1. From the table, it can be seen that the AUC values of all five methods are larger than 0.5, implying that they all have certain ability in monitoring the JIA disease status. Among the five methods, the charts QX-CUSUM, QX-EWMA and NP-EWMA perform better than the charts LQ-EWMA and LQ-CUSUM, and NP-EWMA performs slightly worse than QX-EWMA. The reasoning behind this phenomenon has been demonstrated in You and Qiu (2018) that the data decorrelation involved in the charts LQ-EWMA and LQ-CUSUM would attenuate a potential shift in the observed data and consequently reduce the effectiveness of the charts. The ROC curves of the five charts QX-CUSUM, LQ-CUSUM, QXEWMA, LQ-EWMA and NP-EWMA are shown in Figure 2. From the plots in the figure, it can be seen that QX-CUSUM, QX-EWMA and NP-EWMA have an overall better performance than LQ-EWMA and LQ-CUSUM in this example, although the difference between their ROC curves is quite small. This result is consistent with those in Table 1.

Next, we consider using some multivariate DySS procedures discussed in Section 2, because the original HRQOL data contain a total of 9 PedsQL measures, with 4 in the generic core scales and

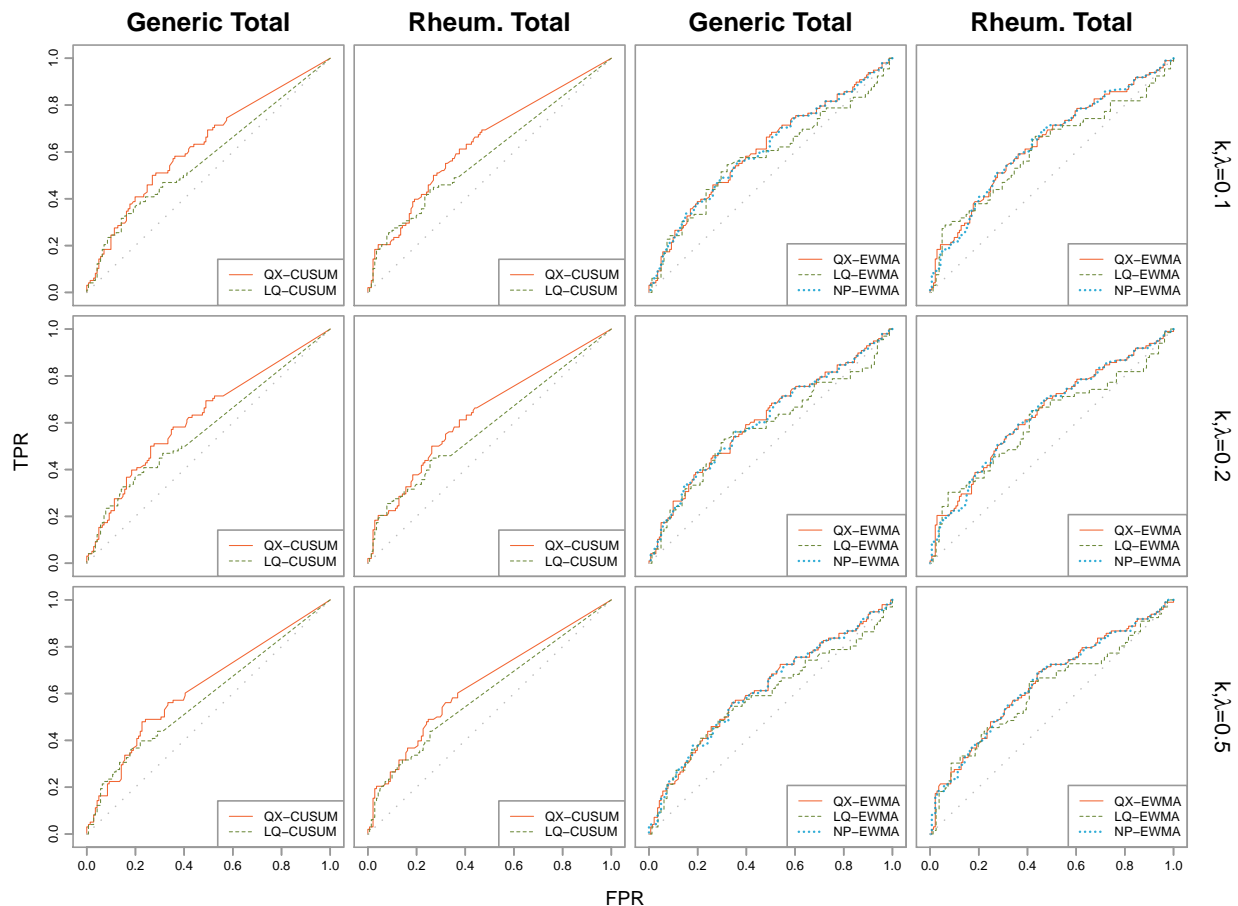


Figure 2: ROC curves of the charts QX-CUSUM, LQ-CUSUM, QX-EWMA, LQ-EWMA, and NP-EWMA when $\lambda = k = 0.1, 0.2$ or 0.5 for monitoring the JIA disease status from the generic total scores or the rheumatology total scores. The dotted diagonal lines denote the ROC curves of random guessing.

5 in the rheumatology module. Before these procedures are applied to all observed data, we should first consider removing some PedsQL measures that are minimally associated with the disease condition of JIA patients. The main reason why we perform this variable selection analysis is that it would not lose much useful information for monitoring the JIA disease condition by removing the PedsQL measures that are minimally associated with the disease condition, and that it can reduce variability in the observed data and make online monitoring of the observed data more effective. For the purpose of variable selection, half of the 126 JIA patients with HDA disease condition are randomly selected and combined with the 171 JIA patients with LDA disease condition in the

IC data to form training data. Then, variable selection is performed using the training data. To this end, let $z_i(t_{ij})$ be a binary indicator, with $z_i(t_{ij}) = 1$ denoting the i th patient in the training data being in HDA disease condition at time t_{ij} , and $z_i(t_{ij}) = 0$ otherwise. Then, we consider the mixed-effect logistic regression model

$$\text{logit} [P(z_i(t_{ij}) = 1)] = \beta_0 + \boldsymbol{\beta}' \mathbf{y}_i(t_{ij}) + \alpha_i,$$

where β_0 is the intercept term, $\boldsymbol{\beta}$ is the q -dimensional regression coefficient, $\text{logit}\{x\} = \log\{x/(1-x)\}$ is the logit link function, and $\alpha_i \sim N(0, \tau^2)$ is the random-effect term. Let $l(\beta_0, \boldsymbol{\beta}, \tau)$ be the log likelihood function, and q_0 be the number of non-zero elements in $\boldsymbol{\beta}$. Then, the Akaike information criterion (AIC) of the model is defined as $-2l(\beta_0, \boldsymbol{\beta}, \tau) + 2(q_0 + 2)$, which is also the default variable selection criterion in the R package *lme4*. By the backward elimination procedure using AIC as the variable selection criterion, the results are shown in Table 2. From the table, it can be seen that the final model selected by this variable selection procedure is obtained after Step 6 with the minimum AIC value of 136.388. This final model contains three measures in the PedsQL rheumatology module: “pain and hurt”, “daily activities”, and “treatment.” This result is consistent with the findings in Varni et al. (2002b) that “pain and hurt” and “daily activities” were consistent with the JIA disease symptoms as causal indicators of HRQOL, and that “pain and hurt” was most responsive to JIA clinical changes.

To check whether the selected final model discussed above would provide the best prediction of the JIA disease condition, we consider using different numbers of PedsQL measures in the multivariate DySS methods. For a given number of PedsQL measures, the variables considered will be based on the ordering provided by Table 2. For instance, if only two PedsQL measures are considered, then we will use the variables “pain and hurt” and “daily activities.” If three PedsQL measures are considered, then “pain and hurt”, “daily activities”, and “treatment” are chosen, and so forth. This is reasonable because Table 2 presents the results of the backward elimination variable selection procedure, the variables are deleted sequentially according to their association with the response, and the variables less associated with the response would be deleted earlier. The multivariate DySS methods considered here include the chart (8)-(9) (denoted as QX-2S), the chart

Table 2: The removed HRQOL measure in each step of the backward elimination variable selection procedure, along with the AIC value of the resulting model. The variables selected by the variable selection procedure are marked by asterisks. The smallest AIC value is displayed in bold.

Step	Removed variable	AIC after removal
1	Social functioning sub-scale	143.682
2	Physical functioning sub-scale	142.029
3	Communication sub-scale	140.442
4	Worry sub-scale	139.201
5	Emotional functioning sub-scale	137.927
6	School functioning sub-scale	136.388
7	Treatment sub-scale*	137.650
8	Daily activities sub-scale*	144.103
9	Pain and hurt sub-scale*	174.393

(10) (denoted as QX-1S), their respective version based on data de-correlation by Li and Qiu (2017) (denoted as LQ-2S and LQ-1S), the method by simultaneously monitoring individual variables using the univariate EWMA chart (2) (denoted as SIMUL), the method by simultaneously monitoring individual variables using the nonparametric EWMA chart (7) (denoted as NP-SIMUL), and the one-sided multivariate CUSUM chart (11) (denoted as MCUSUM). In each of these charts, the weighting parameter λ or the allowance constant k is chosen to be $\lambda = k = 0.1, 0.2$ and 0.5 . The results of their AUC values, computed by the bootstrap procedure with the bootstrap sample size of 1000, are presented in Table 3. From the table, we can have the following conclusions. First, LQ-2S and LQ-1S are less effective than QX-2S and QX-1S, which is consistent with the results of Table 1 in univariate cases. Second, QX-1S is more effective than QX-2S. This is reasonable because QX-1S is designed for detecting downward shifts only while QX-2S is designed for detecting arbitrary shifts. In the current JIA example, OC patients have relatively low PedsQL measures, compared to the IC patients, as shown in Figure 1. Third, it seems that SIMUL, NP-SIMUL and MCUSUM perform well in this example, and MCUSUM is slightly better than the other two methods. Note that SIMUL, NP-SIMUL and MCUSUM are designed for detecting downward shifts. SIMUL and NP-SIMUL ignore the association among individual variables during sequential monitoring, while

MCUSUM takes into account the association in its charting statistic, which explains why MCUSUM is a bit more effective. Fourth, QX-1S, SIMUL, NP-SIMUL and MCUSUM perform the best when the selected final model discussed in Table 2 is used (i.e., “dim=3” in Table 3). So, it seems that it is helpful for detecting HDA in JIA by deleting certain PedsQL measures that are minimally associated with the JIA disease conditions. Also, by comparing the results in Tables 1 and 3, it can be seen that it is generally more effective to use multiple PedsQL measures to detect HDA in JIA.

To further demonstrate the benefits of removing certain predictors when monitoring the JIA disease status, Figure 3 shows the ROC curves of the four charts QX-1S, SIMUL, NP-SIMUL, and MCUSUM, which have relatively good results in Table 3, in cases when they use all 9 PedsQL measures and when they use the 3 PedsQL measures “pain and hurt”, “daily activities”, and “treatment” in the final selected model only. It can be seen from the plots of Figure 3 that the performance of the four charts is generally improved after the 6 PedsQL measures that are minimally associated with the JIA disease status are removed from sequential monitoring.

4 Simulation Studies

In the previous section, the performance of different multivariate DySS methods are evaluated using the JIA data. To further compare their performance, we present some simulation results in this section. In the simulation study, $q = 10$ dimensions of patient-reported outcomes (e.g., HRQOL scores in the JIA example) are sequentially collected for n patients. The outcomes of the i th patient, denoted as $\mathbf{y}_i(t) = (y_{i,1}(t), \dots, y_{i,q}(t))'$, are assumed to follow the longitudinal model $\mathbf{y}_i(t) = \boldsymbol{\mu}(t) + \boldsymbol{\epsilon}_i(t)$, for $t \in [0, T]$, where $\boldsymbol{\mu}(t) = (\mu_1(t), \dots, \mu_q(t))'$, and $\mu_j(t) = \log(1 + 0.001t)$ for each j . The error term $\boldsymbol{\epsilon}_i(t)$ is assumed to be $\boldsymbol{\epsilon}_i(t) = \mathbf{v}_i + \boldsymbol{\epsilon}_{i,0}(t)$, where \mathbf{v}_i is a time-independent random effect term that introduces both between-subject and within-subject correlation, and $\boldsymbol{\epsilon}_{i,0}(t)$ can be interpreted as the pure measurement error. To be more specific, let \mathbf{v}_i follow a multivariate

Table 3: Bootstrap estimates of AUC values and their standard errors (in parentheses) of different multivariate DySS methods. Dim denotes the number of variables considered in the multivariate DySS methods. Numbers in bold denote the largest numbers in individual columns.

Dim	QX-2S	LQ-2S	QX-1S	LQ-1S	SIMUL	NP-SIMUL	MCUSUM
$\lambda, k = 0.1$							
2	0.5968 (0.0017)	0.5438 (0.0013)	0.6310 (0.0017)	0.5755 (0.0014)	0.6403 (0.0017)	0.6308 (0.0016)	0.6411 (0.0016)
3	0.5912 (0.0018)	0.5425 (0.0013)	0.6346 (0.0015)	0.5597 (0.0017)	0.6487 (0.0015)	0.6338 (0.0015)	0.6543 (0.0014)
4	0.6135 (0.0018)	0.5513 (0.0015)	0.6234 (0.0017)	0.5674 (0.0016)	0.6299 (0.0016)	0.6202 (0.0016)	0.6403 (0.0015)
5	0.5984 (0.0021)	0.5458 (0.0021)	0.5761 (0.0021)	0.5547 (0.0018)	0.5919 (0.0017)	0.5938 (0.0017)	0.6081 (0.0017)
6	0.6031 (0.0020)	0.5918 (0.0022)	0.6050 (0.0018)	0.5608 (0.0020)	0.6081 (0.0017)	0.5931 (0.0017)	0.6198 (0.0018)
7	0.5777 (0.0015)	0.5456 (0.0025)	0.6142 (0.0017)	0.5370 (0.0020)	0.6121 (0.0017)	0.6006 (0.0017)	0.6288 (0.0017)
8	0.5814 (0.0015)	0.5279 (0.0022)	0.6053 (0.0016)	0.5205 (0.0017)	0.6103 (0.0017)	0.5974 (0.0017)	0.6256 (0.0016)
9	0.5828 (0.0017)	0.5354 (0.0021)	0.5864 (0.0018)	0.5290 (0.0013)	0.5973 (0.0017)	0.5961 (0.0016)	0.6071 (0.0020)
$\lambda, k = 0.2$							
2	0.5947 (0.0017)	0.5413 (0.0014)	0.6338 (0.0017)	0.5769 (0.0014)	0.6432 (0.0016)	0.6306 (0.0017)	0.6419 (0.0016)
3	0.5859 (0.0017)	0.5435 (0.0013)	0.6348 (0.0015)	0.5609 (0.0017)	0.6514 (0.0015)	0.6352 (0.0015)	0.6538 (0.0014)
4	0.6091 (0.0017)	0.5505 (0.0016)	0.6238 (0.0017)	0.5700 (0.0017)	0.6335 (0.0016)	0.6221 (0.0017)	0.6417 (0.0015)
5	0.5922 (0.0023)	0.5491 (0.0019)	0.5742 (0.0023)	0.5522 (0.0018)	0.5967 (0.0017)	0.5947 (0.0017)	0.6079 (0.0018)
6	0.6035 (0.0020)	0.5887 (0.0022)	0.6057 (0.0019)	0.5613 (0.0020)	0.6115 (0.0018)	0.5943 (0.0017)	0.6176 (0.0019)
7	0.5777 (0.0015)	0.5437 (0.0025)	0.6178 (0.0016)	0.5372 (0.0020)	0.6136 (0.0017)	0.6016 (0.0017)	0.6257 (0.0018)
8	0.5858 (0.0015)	0.5277 (0.0022)	0.6073 (0.0016)	0.5208 (0.0017)	0.6107 (0.0018)	0.5986 (0.0017)	0.6241 (0.0016)
9	0.5870 (0.0016)	0.5354 (0.0021)	0.5878 (0.0018)	0.5294 (0.0013)	0.5987 (0.0017)	0.5947 (0.0016)	0.6064 (0.0020)
$\lambda, k = 0.5$							
2	0.5856 (0.0017)	0.5460 (0.0014)	0.6325 (0.0016)	0.5842 (0.0015)	0.6394 (0.0016)	0.6314 (0.0018)	0.6379 (0.0016)
3	0.5920 (0.0016)	0.5562 (0.0014)	0.6348 (0.0015)	0.5681 (0.0018)	0.6534 (0.0015)	0.6451 (0.0015)	0.6539 (0.0014)
4	0.6094 (0.0019)	0.5558 (0.0016)	0.6211 (0.0019)	0.5707 (0.0017)	0.6405 (0.0015)	0.6322 (0.0015)	0.6439 (0.0015)
5	0.6190 (0.0017)	0.5571 (0.0016)	0.5758 (0.0026)	0.5606 (0.0014)	0.6086 (0.0017)	0.6081 (0.0016)	0.6113 (0.0018)
6	0.6220 (0.0017)	0.5870 (0.0021)	0.6025 (0.0023)	0.5622 (0.0018)	0.6209 (0.0017)	0.6083 (0.0017)	0.6170 (0.0020)
7	0.5772 (0.0016)	0.5416 (0.0023)	0.6165 (0.0017)	0.5340 (0.0020)	0.6219 (0.0016)	0.6146 (0.0016)	0.6266 (0.0018)
8	0.5824 (0.0015)	0.5288 (0.0021)	0.6051 (0.0017)	0.5205 (0.0016)	0.6171 (0.0018)	0.6112 (0.0016)	0.6238 (0.0017)
9	0.5757 (0.0016)	0.5359 (0.0019)	0.5831 (0.0019)	0.5304 (0.0012)	0.5997 (0.0018)	0.6027 (0.0016)	0.6055 (0.0020)

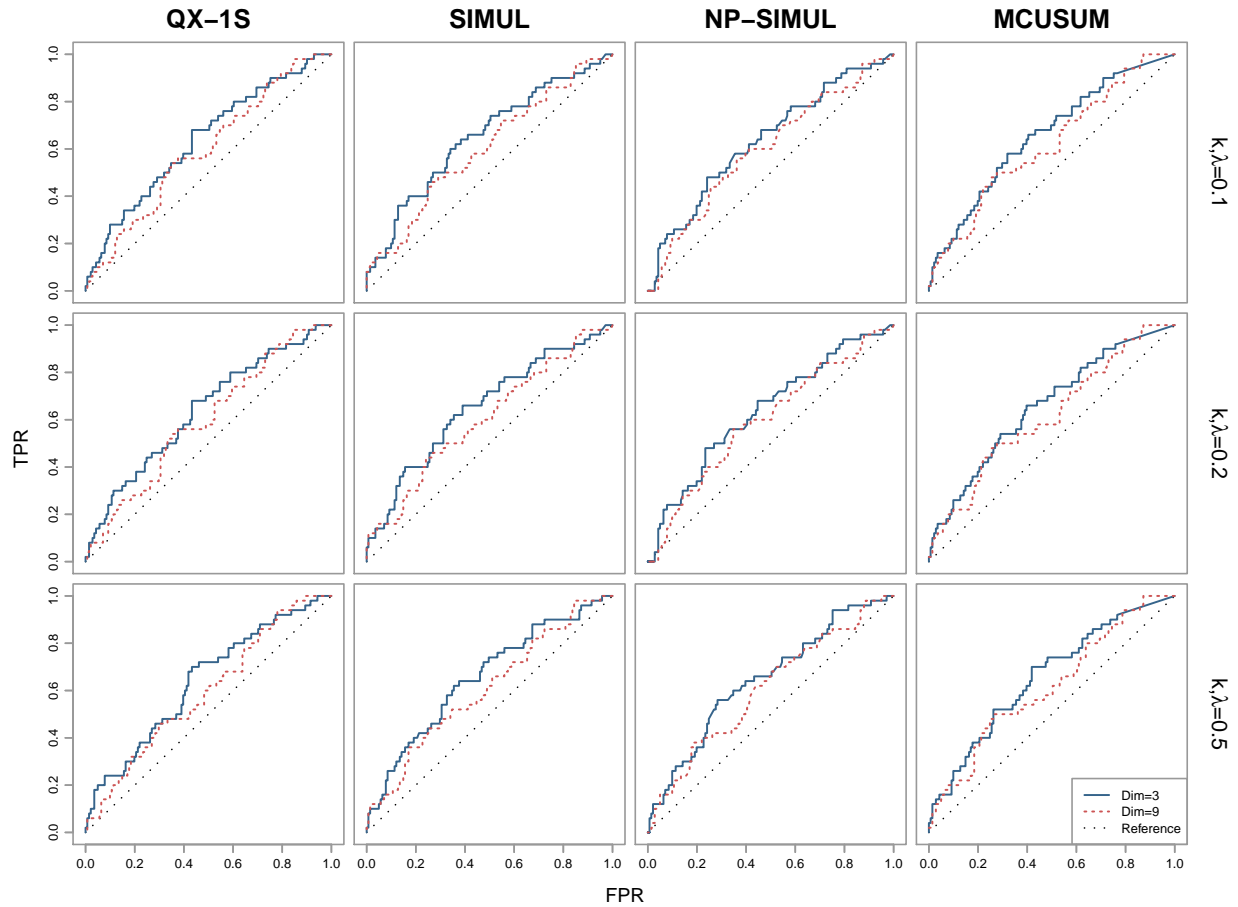


Figure 3: ROC curves of the charts QX-1S, SIMUL, NP-SIMUL, and MCUSUM when $\lambda = k = 0.1$, 0.2 or 0.5 for monitoring the JIA disease status from all 9 PedsQL measures (dashed curves, Dim=9) or from the 3 PedsQL measures “pain and hurt”, “daily activities”, and “treatment” (solid curves, Dim=3) in the final selected model of the model selection procedure (cf., Table 2). The dotted diagonal lines denote the ROC curves of random guessing, presented as references.

normal distribution with mean $\mathbf{0}$ and covariance matrix

$$\begin{pmatrix} 0.5 & 0.2 & \cdots & 0.2 \\ 0.2 & 0.5 & \cdots & 0.2 \\ \vdots & \vdots & \ddots & \vdots \\ 0.2 & 0.2 & \vdots & 0.5 \end{pmatrix},$$

and $\boldsymbol{\epsilon}_{i,0}(t)$ be realizations of independent normal random vectors with mean $\mathbf{0}$ and covariance matrix $\mathbf{I}_{q \times q}$. The disease activity scores are assumed to be associated with the first 3 patient-reported outcome variables through the model

$$\xi_i(t_{ij}) = -y_{i,1}(t_{ij}) - y_{i,2}(t_{ij}) - y_{i,3}(t_{ij}) + \tilde{\epsilon}_i(t_{ij}),$$

where $\xi_i(t_{ij})$ denotes the observed disease activity score of the i th patient at t_{ij} , and $\tilde{\epsilon}_i(t_{ij})$ is an error term. Let γ be the cutoff value for defining disease activity status. Namely, a patient is said to be in high disease activity if $\xi_i(t_{ij}) \geq \gamma$, and in low disease activity otherwise. The observation times are generated by $t_{i,j} - t_{i,j-1} = \Delta_{i,j}$, for $i = 1, \dots, n$ and $j = 1, \dots, m_i$, where $t_{i,0} = 0$, $\{\Delta_{i,j}\}$ are generated independently from the Poisson distribution with the mean of 100. All observation times are assumed to be in the design interval $[0, 3000]$. Here, 3000 is chosen because it is close to the length of 3395 of the study period in the JIA data example. Similarly, the mean value 100 of $\{\Delta_{i,j}\}$ is chosen to be close to the mean distance of 78.6 days between two consecutive observation times in the JIA data example.

In the simulation study, the sample size n is chosen to be 100 or 200, the weighting parameter λ of the EWMA charts is chosen to be 0.1, 0.2 or 0.5, the allowance constant k of the CUSUM charts is chosen to be 0.1, 0.2 or 0.5, and the cutoff value γ for defining activity status is chosen to be the 50th or 75th percentiles of $\{\xi_i(t_{ij})\}$. The variable selection procedure described in Section 3.2 is used to delete certain outcome variables before online monitoring. All DySS methods considered are used for monitoring 1,000 simulated new patients. All simulations are repeated for 100 times. The AUC values of different methods are presented in Table 4. In the table, ‘‘Full’’ represents the case where all the q patient-reported outcomes are monitored, while ‘‘Reduced’’ represents the case where only the ones selected by the variable selection procedure are monitored. From the table, it

can be seen that i) variable selection indeed improves the performance of DySS methods in terms of AUC in all cases considered, ii) one-sided control charts out-perform their two-sided counterparts in most cases considered, iii) QX-1S, SIMUL, NP-SIMUL and MCUSUM still demonstrate good performance in most cases considered, and iv) MCUSUM is a little bit more efficient than the others in terms of AUC. All these results are consistent with the those in the JIA example discussed in Section 3.

5 Concluding Remarks

JIA in its HDA stage can seriously affect patients' physical and psychological health. Because no effective cure of JIA has been found, it is important to detect the HDA stage for a JIA patient early so that proper disease management and treatment can be made in a timely fashion to alleviate the disease symptoms and improve patients' overall quality of life. To this end, the traditional approach based on the JADAS scores is often time consuming, costly, and inconvenient to patients. Instead, PedsQL measures are much more convenient to obtain. In the previous sections, we have shown that the recent statistical methodology DySS is effective to detect the HDA disease status for JIA patients by sequentially monitoring the PedsQL measures of a JIA patient. To improve the performance of a DySS procedure, we suggested removing certain PedsQL measures that are minimally associated with the JIA disease conditions by a variable selection procedure. Among several different versions of DySS, QX-1S, SIMUL, NP-SIMUL and MCUSUM are recommended, among which MCUSUM seems to have a little bit better performance. There still are some issues that need to be addressed in our future research. For instance, in variable selection, we delete a variable completely if the resulting model has a smaller AIC value. An alternative approach is to use a weighted average of all variables with the ones less associated with the JIA disease conditions receiving less weights. It is unknown to us whether this alternative approach would be more effective. Also, the current paper evaluates the performance of a DySS procedure using the false positive rate and the true positive rate in terms of the area under the ROC curve. Some alternative performance measures (e.g., the trimmed area under the ROC curve) are possible. It is

Table 4: AUC values and their standard errors (in parantheses) of different multivariate DySS methods in the simulation study. “Full” represents the case where all the q patient-reported outcomes are monitored, while “Reduced” represents the case where only the ones selected by the variable selection procedure are monitored.

n	k, λ	Model	QX-2S	LQ-2S	QX-1S	LQ-1S	SIMUL	NP-EWMA	MCUSUM
0.50 percentile									
100	0.1	Full	0.708 (0.002)	0.568 (0.003)	0.790 (0.001)	0.608 (0.003)	0.815 (0.001)	0.807 (0.001)	0.854 (0.001)
		Reduced	0.761 (0.003)	0.741 (0.004)	0.842 (0.002)	0.754 (0.005)	0.850 (0.001)	0.850 (0.001)	0.881 (0.001)
	0.2	Full	0.706 (0.002)	0.567 (0.003)	0.796 (0.001)	0.600 (0.003)	0.821 (0.001)	0.814 (0.001)	0.854 (0.001)
		Reduced	0.760 (0.003)	0.734 (0.003)	0.841 (0.001)	0.742 (0.004)	0.849 (0.001)	0.849 (0.001)	0.881 (0.001)
	0.5	Full	0.697 (0.002)	0.562 (0.002)	0.795 (0.001)	0.594 (0.003)	0.821 (0.001)	0.819 (0.001)	0.854 (0.001)
		Reduced	0.744 (0.003)	0.686 (0.003)	0.834 (0.001)	0.703 (0.004)	0.843 (0.001)	0.843 (0.001)	0.881 (0.001)
200	0.1	Full	0.710 (0.002)	0.563 (0.002)	0.791 (0.001)	0.616 (0.003)	0.816 (0.000)	0.807 (0.001)	0.854 (0.000)
		Reduced	0.763 (0.003)	0.694 (0.003)	0.841 (0.002)	0.727 (0.003)	0.849 (0.001)	0.849 (0.001)	0.881 (0.001)
	0.2	Full	0.709 (0.002)	0.563 (0.002)	0.797 (0.001)	0.609 (0.002)	0.822 (0.000)	0.814 (0.000)	0.854 (0.000)
		Reduced	0.762 (0.003)	0.691 (0.003)	0.841 (0.001)	0.717 (0.003)	0.849 (0.001)	0.849 (0.001)	0.881 (0.001)
	0.5	Full	0.700 (0.002)	0.553 (0.002)	0.797 (0.001)	0.597 (0.002)	0.822 (0.001)	0.819 (0.000)	0.854 (0.000)
		Reduced	0.746 (0.002)	0.649 (0.002)	0.834 (0.001)	0.683 (0.003)	0.843 (0.001)	0.843 (0.001)	0.880 (0.001)
0.75 percentile									
100	0.1	Full	0.705 (0.002)	0.544 (0.002)	0.805 (0.001)	0.576 (0.003)	0.837 (0.001)	0.833 (0.000)	0.874 (0.000)
		Reduced	0.751 (0.003)	0.702 (0.003)	0.845 (0.001)	0.711 (0.004)	0.857 (0.001)	0.858 (0.001)	0.890 (0.001)
	0.2	Full	0.698 (0.002)	0.543 (0.002)	0.805 (0.001)	0.558 (0.003)	0.838 (0.000)	0.837 (0.000)	0.874 (0.000)
		Reduced	0.747 (0.003)	0.704 (0.003)	0.840 (0.001)	0.697 (0.004)	0.853 (0.001)	0.855 (0.001)	0.890 (0.001)
	0.5	Full	0.690 (0.002)	0.544 (0.002)	0.796 (0.001)	0.560 (0.002)	0.831 (0.001)	0.833 (0.001)	0.874 (0.000)
		Reduced	0.730 (0.003)	0.674 (0.003)	0.836 (0.001)	0.677 (0.003)	0.850 (0.001)	0.852 (0.001)	0.890 (0.001)
200	0.1	Full	0.707 (0.002)	0.538 (0.002)	0.807 (0.001)	0.581 (0.002)	0.838 (0.000)	0.834 (0.000)	0.875 (0.000)
		Reduced	0.752 (0.002)	0.669 (0.003)	0.846 (0.001)	0.691 (0.003)	0.858 (0.001)	0.859 (0.001)	0.890 (0.001)
	0.2	Full	0.700 (0.002)	0.538 (0.002)	0.806 (0.001)	0.561 (0.002)	0.839 (0.000)	0.838 (0.000)	0.875 (0.000)
		Reduced	0.748 (0.002)	0.674 (0.003)	0.841 (0.001)	0.679 (0.003)	0.854 (0.001)	0.855 (0.001)	0.890 (0.001)
	0.5	Full	0.693 (0.002)	0.537 (0.001)	0.797 (0.001)	0.560 (0.002)	0.833 (0.000)	0.834 (0.000)	0.874 (0.000)
		Reduced	0.731 (0.002)	0.653 (0.002)	0.836 (0.001)	0.664 (0.003)	0.851 (0.001)	0.852 (0.001)	0.890 (0.001)

worth studying the performance of different DySS procedures under such alternative performance measures. The CUSUM and EWMA control charts used in the DySS procedures depend on the parameters k and λ . In the SPC literature, it is well known that smaller values of k and λ are good for detecting smaller shifts while larger values of k and λ are good for detecting larger shifts. However, in the JIA application, it is usually unknown to us how the HRQOL scores of the JIA patients at the HDA status would deviate from the mean scores of the JIA patients at the LDA status. Thus, selection of these parameters could be difficult. Several existing methods to overcome this difficulty will be considered in our future research. For example, Sparks (2001) proposed an adaptive CUSUM chart that continually adjusted the parameter k after the shift size is estimated at each observation time point. Qiu et al. (2018) proposed a hybrid method by combining a set of admissible values for the parameter λ in the EWMA chart.

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

References

- Consolaro, A., Giancane, G., Schiappapietra, B., Davì, S., Calandra, S., Lanni, S., and Ravelli, A. (2016). “Clinical outcome measures in juvenile idiopathic arthritis,” *Pediatric Rheumatology* **14**, 23.
- Consolaro, A., Ruperto, N., Bazso, A., Pistorio, A., Magni-Manzoni, S., Filocamo, G., Malattia, C., Viola, S., Martini, A., and Ravelli, A. (2009). “Development and validation of a composite disease activity score for juvenile idiopathic arthritis,” *Arthritis Care & Research: Official Journal of the American College of Rheumatology* **61**, 658–666.
- Consolaro, A., Van Dijkhuizen, P., Espada, G., Varbanova, B., Oliveira-Knupp, S., Miettunen, P., Chedeville, G., Hofer, M., Dolezalova, P., Foeldvari, I., et al. (2017). “Development of new juvenile arthritis disease activity score cut-offs for oligoarthritis and RF-negative

- polyarthritis from a large multinational cohort of children with juvenile idiopathic arthritis,” *2017 ACR/ARHP Annual Meeting*.
- Crosier, R. B. (1986). “A new two-sided cumulative sum quality control scheme,” *Technometrics* **28**, 187–194.
- Fricker, R. D. (2007). “Directionally sensitive multivariate statistical process control methods with application to syndromic surveillance,” *International Society for Disease Surveillance* **3**.
- Giancane, G., Consolaro, A., Lanni, S., Davì, S., Schiappapietra, B., and Ravelli, A. (2016). “Juvenile idiopathic arthritis: diagnosis and treatment,” *Rheumatology and Therapy* **3**, 187–207.
- 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.
- Mańczak, M., Rutkowska-Sak, L., and Raciborski, F. (2016). “Health-related quality of life in children with juvenile idiopathic arthritis – child’s and parent’s point of view,” *Reumatologia* **54**, 243.
- McErlane, F., Beresford, M. W., Baidam, E. M., Chieng, S. A., Davidson, J. E., Foster, H. E., Gardner-Medwin, J., Lunt, M., Wedderburn, L. R., Thomson, W., et al. (2013). “Validity of a three-variable juvenile arthritis disease activity score in children with new-onset juvenile idiopathic arthritis,” *Annals of the Rheumatic Diseases* **72**, 1983–1988.
- Moustakides, G. V. (1986). “Optimal stopping times for detecting changes in distributions,” *Annals of Statistics* **14**, 1379–1387.
- Pepe, M. S. (2003). *The Statistical Evaluation of Medical Tests for Classification and Prediction*, Oxford: 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 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.
- Seid, M., Huang, B., Niehaus, S., Brunner, H. I., and Lovell, D. J. (2014). “Determinants of health-related quality of life in children newly diagnosed with juvenile idiopathic arthritis,” *Arthritis Care & Research* **66**, 263–269.
- Seid, M., Opiari, L., Huang, B., Brunner, H. I., and Lovell, D. J. (2009). “Disease control and health-related quality of life in juvenile idiopathic arthritis,” *Arthritis Care & Research* **61**, 393–399.
- Sparks, R. S. (2000). “CUSUM charts for signalling varying location shifts,” *Journal of Quality Technology*, **32**, 157–171.
- Varni, J. W., Seid, M., Knight, T. S., Uzark, K., and Szer, I. S. (2002a). “The PedsQL™ 4.0 Generic Core Scales: Sensitivity, responsiveness, and impact on clinical decision-making,” *Journal of Behavioral Medicine* **25**, 175–193.
- Varni, J. W., Seid, M., Smith Knight, T., Burwinkle, T., Brown, J., and Szer, I. S. (2002b). “The PedsQL™ in pediatric rheumatology: reliability, validity, and responsiveness of the Pediatric Quality of Life Inventory™ Generic Core Scales and Rheumatology Module,” *Arthritis & Rheumatism* **46**, 714–725.

- Weiss, P. F., Klink, A. J., Faerber, J., and Feudtner, C. (2013). “The pediatric rheumatology quality of life scale: validation of the English version in a us cohort of juvenile idiopathic arthritis,” *Pediatric Rheumatology* **11**, 43.
- WHOQOL Group. (1998). “Development of the World Health Organization WHOQOL-BREF quality of life assessment,” *Psychological medicine* **28**, 551–558.
- Woodall, W. H. and Ncube, M. M. (1985). “Multivariate cusum quality-control procedures,” *Technometrics* **27**, 285–292.
- Wu, Q., Chaplin, H., Ambrose, N., Sen, D., Leandro, M. J., Wing, C., Daly, N., Webb, K., Fisher, C., Suffield, L., et al. (2016). “Juvenile arthritis disease activity score is a better reflector of active disease than the disease activity score 28 in adults with polyarticular juvenile idiopathic arthritis,” *Annals of the Rheumatic Diseases* **75**, 635–636.
- Xiang, D., Qiu, P., and Pu, X. (2013) “Nonparametric regression analysis of multivariate longitudinal data,” *Statistica Sinica* **23**, 769–789.
- You, L., and Qiu, P. (2019). “Fast computation for dynamic screening systems when observations are correlated,” *Journal of Statistical Computation and Simulation* **89**, 379–394.