Statistica Sinica **13**(2003), 593-612

AUC-BASED TESTS FOR NONPARAMETRIC FUNCTIONS WITH LONGITUDINAL DATA

Yanqing Sun and Hulin Wu

University of North Carolina and University of Rochester

Abstract: Longitudinal data are very common in biomedical and clinical research, for example, CD4+ cell responses and viral load responses in AIDS clinical research. It is challenging to do inference for the whole trajectory of these longitudinal data if a parametric function is not available to model the trajectories. In this paper we develop an area-under-the-curve (AUC) based nonparametric method to compare the two groups of longitudinal data under both fixed and random designs. The proposed test does not involve any smoothing. The method is also applicable to one-sample problems. The test statistic is based on the maximum deviation of the weighted averages of AUCs between two groups. The weight functions are used to account for censored or early drop-out subjects. For both cases that the number of measurements per subject goes to infinity and is finite, we show that the test statistic processes converge weakly to Gaussian processes, where for the case of the number of measurements per subject going to infinity, a nonparametric mixed-effects model is considered. A Monte Carlo method is developed to generate the distribution of test statistics. Simulations show that the test is valid and promising. We applied the test to compare CD4+ responses over time between two treatment groups in an AIDS clinical trial.

Key words and phrases: Censoring, confidence bands, fixed and random designs, nonparametric maximum deviation tests, nonparametric mixed-effects, one and two-sample problems.

1. Introduction

Longitudinal data arise frequently in biomedical and clinical research. Linear and nonlinear parametric models for longitudinal data have been intensively studied in the past two decades. Good surveys can be found in the books by Diggle, Liang and Zeger (1994), Davidian and Giltinan (1995) and Vonesh and Chinchilli (1996). Recently nonparametric and semiparametric models for longitudinal data have been paid a great attention due to the needs of scientific research and the lack of development in this area. Among others, these works include semiparametric methods by Moyeed and Diggle (1994) and Zeger and Diggle (1994), nonparametric kernel and spline methods by Hoover, Rice, Wu and Yang (1998), Wu, Chiang and Hoover (1998) and Scheike and Zhang (1998).

Martinussen and Scheike (1999, 2000, 2001) and Lin and Ying (2001) have developed complete nonparametric methods based on the estimates of the cumulative regression functions without involving any smoothing.

In this paper, we propose a test based on the concept of area under the curve (AUC) to compare the longitudinal responses between two treatment groups. The development of the methodology is motivated by an AIDS clinical study in which the investigators want to compare the immune responses (measured by CD4+ cell counts) in HIV-1 infected children between two treatment groups. Children in cohort 1 were treated with a lower dose antiviral therapy while children in cohort 2 were treated with a higher dose of the same therapy. We plot the changes in CD4+ cell counts from baseline to 48 weeks for all subjects in Figure 1 (a). Note that it is difficult to model the CD4+ cell response by a parametric function and the CD4+ cell counts for each patient are correlated over time. A nonparametric longitudinal model is a natural choice.

Consider a random sample of n subjects. For the ith subject, $i = 1, \ldots, n$, let $Y_i(t)$ be the response variable at time t. The response curves follow the nonparametric regression model

$$Y_i(t) = \eta(t) + \epsilon_i(t), \quad a \le t \le b,\ i = 1, \ldots, n, \tag{1.1}$$

where $\eta(t)$ is the mean response curve and $\epsilon_i(t)$ is a zero-mean stochastic process. The response curve for subject i is observed at times $\{t_{ij}^o, j = 1, \ldots, n_i\}$. The area under the mean curve is defined as $\int_a^t \eta(s)\, ds$ for each $t \in [a, b]$. The concept of AUC has been widely used in pharmacokinetics/pharmacodynamics where the AUC is the area under the drug-concentration-time curve and is used to measure the total drug exposure in the body of a patient (Rowland and Tozer (1995)). The AUC of HIV-1 RNA copies (viral load) was also used to measure the cumulative effect of a response in AIDS clinical trials (Weinberg and Lagakos (2000)).

Scheike and Zhang (1998), Scheike, Zhang and Juul (1999) and Scheike (2000) have proposed use of the cumulative function to compare two nonparametric functions. The proposed method for longitudinal data in Scheike and Zhang (1998) involves smoothing of two groups of curves. Thus, the test is sensitive to the degree of smoothing or smoothing parameters. Scheike (2000) suggested using the cumulative Priestley-Chao estimator to construct the test for comparing two curves for i.i.d. data. Martinussen and Scheike (1999, 2000, 2001) and Lin and Ying (2001) recently considered the varying-coefficient regression models for longitudinal data and have shown that the cumulative function is much easier to estimate, and more efficient for inferences, than the regression function itself.

We propose an AUC based nonparametric method to compare the two groups of longitudinal data under both fixed and random designs. The test statistic is based on the maximum deviation of the weighted averages of the AUCs between the two groups. The weight functions account for individuals who are censored or drop out early from the study. To derive the asymptotic distribution of the test statistics, we consider two cases. One case is that the number of measurements per subject goes to infinity where a nonparametric mixed-effects model is considered, and another is the case of a finite number of measurements per subject. For both cases, we show that the test statistic processes converge weakly to Gaussian processes. A Monte Carlo method is developed to generate the distribution of test statistics. Our method does not involve any smoothing and a parametric function is not assumed for the underlying response variable. Simulations show that the test is valid and promising. The difference between the proposed method and the aforementioned methods derived from the estimates of the cumulative functions is that our tests are constructed from the AUCs of all subjects, which brings in the subject-specific feature of longitudinal data. Further, we study the case where the number of measurements per subject goes to infinity, in addition to the finite number of measurements per subject, see Hoover and Wu (2001). Our method is also applicable to one-sample problems. Fortran subroutines and Splus functions for implementing the tests are ready to use.

The paper is organized as follows. Section 2 contains a description of an individual area under the curve and the random process constructed from the weighted individual AUCs. The weak convergence of the random process is derived. In Section 3, we propose tests for both one and two sample cases. The AUC-based confidence bands are also discussed. Relevant asymptotic results are presented. Section 4 presents some simulation results for both fixed and random design, and we illustrate the proposed method by an application example from an AIDS clinical study for comparing CD4+ cell counts between two treatment groups. All proofs are relegated to Section 5.

2. Area Under the Curve (AUC)

In model (1.1), let $\{Y_i(t), \epsilon_i(t), i = 1, \ldots, n\}$ be independent identically distributed (i.i.d.) random processes. We consider both random and fixed designs. Let $\{t^o_{ij}, j = 1, \ldots, n_i, i = 1, \ldots, n\}$ be independently distributed according to a density function $f(t), a \leq t \leq b$. Let $t_{i1} < t_{i2} < \ldots < t_{in_i}$ be the ordered values of $\{t^o_{ij}, j = 1, \ldots, n_i\}$ for each i and $t_{i0} = a$ be the entry time of the study. For random designs, the response curve for the subject i is observed at $\{t_{ij}, j = 1, \ldots, n_i\}$ and the observed response values are $y_{ij} = Y_i(t_{ij}), j = 1, \ldots, n_i$. For fixed designs, all subjects are observed at the same time intervals, $\{t_{1j}, j = 1, \ldots, n_1\}$. The number of observations taken on the ith subject by time t is $N_i(t) = \sum_{j=1}^{n_i} I(t_{ij} \leq t)$, where $I(\cdot)$ is the indicator function. In the longitudinal studies, subjects are followed over a period of time and the responses are taken at different time points. It is not surprising that some subjects may drop out of the study early. For example, if a patient under a certain treatment is supposed to visit a clinic monthly to measure CD4+ cell counts for a period of one year, the patient may drop out of the study after a few visits due to loss of follow-up or death. Let C_i be the end of follow-up time or censoring time for the ith subject. Then the responses for the ith subject can only be observed at the time points before C_i. Assume that $\{C_i,\ i = 1, \ldots, n\}$ are independent identically distributed and independent of $\{t_{ij}\}$ and $\{y_{ij}\}$. Throughout the paper, for notational convenience, we drop the index i in the expectation $E\{X_i(t)\}$ when $X_i(t)$'s are identically distributed.

Note that there is a unique relationship between the mean curve function $\eta(t)$ and the AUC function, $\int_0^t \eta(s)\,ds$, $a \le t \le b$. The area under each individual curve over time reflects the sample information of the response curve. One of the simple estimation methods of the area under an individual curve is by the *trapezoidal rule*, which is the area under the curve connecting the observed time and response points through straight line segments (Rowland and Tozer (1995)). The estimation of the area under the curve for subject i between a and t, for $a \le t \le b$, is given by

$$AUC_i(t) = \sum_{j=1}^{N_i(t)} [0.5(y_{ij-1} + y_{ij})(t_{ij} - t_{ij-1})] + y_{iN_i(t)}(t - t_{iN_i(t)}). \tag{2.1}$$

Let $\xi_i(t) = I(C_i \ge t)$. We define the weighted average of AUCs for n subjects as

$$\overline{AUC}(t) = \sum_{i=1}^{n} w_i(t) AUC_i(t), \tag{2.2}$$

where $w_i(t) = \xi_i(t)/\sum_{i=1}^n \xi_i(t)$ is the weight function. Note that $\sum_{i=1}^n w_i(t) = 1$. The $\overline{AUC}(t)$ is the average of areas under the curves that are still not censored at the time t. As t increases, the number of curves that are not censored decreases. Other choices of weight functions are possible, for example, $w_i(t) = \xi_i(t)N_i(t)/\sum_{i=1}^n \xi_i(t)N_i(t)$. Under random design, this weight function puts more weight on curves that have more measurements observed before time t. In our simulation study (not shown here), we found that, when the variation in the numbers of measurements among subjects is large under random designs, the tests using the weight function with $N_i(t)$ are more powerful. For simplicity of theoretical treatment, we only consider $w_i(t) = \xi_i(t)/\sum_{i=1}^n \xi_i(t)$.

It is expected that when the number of the design time points is large, the $\overline{AUC}(t)$ approximates the area under the mean curve $\int_0^t \eta(s)\,ds$ well, while this may not be the case when the number of the design time points is small. We discuss both situations. Consider the nonparametric mixed-effects version of (1.1) with the following error process decomposition:

$$\epsilon_i(t) = v_i(t) + e_i(t), \tag{2.3}$$

where $v_i(t)$ is a mean-zero process satisfying $|v_i(t)-v_i(s)| \le L_i|t-s|$, $Ev_i^4(a) < \infty$, $EL_i^4 < \infty$ and $e_i(t)$ is a white noise process with continuously differentiable variance. Under the nonparametric mixed-effects model, the population mean response is $\eta(t)$ while the subject-specific mean response for individual i is $\eta(t) + v_i(t)$. The random effect $v_i(t)$ also induces correlation among observations within subject i. Examples of (2.3) include the specification $\epsilon(t_{ij}) = v_i + Z_{ij}$ and $\epsilon(t_{ij}) = m(t)v_i + Z_{ij}$, where $\{v_i\}$ are i.i.d. mean zero random effects, $\{Z_{ij}\}$ are i.i.d. mean zero measurement errors and $m(t)$ satisfies a Lipschitz condition. The nonparametric mixed-effects model has been very useful in modelling longitudinal data; see Shi, Weiss and Taylor (1996), Rice and Wu (2001) and Wu and Zhang (2002).

The following large sample properties of $\overline{AUC}(t)$ hold for both random and fixed designs.

Theorem 1. *Suppose $f(t)$ is positive, continuously differentiable and bounded below on $[a,b]$, the distribution of C is continuous with $P(C > b) > 0$, and $\eta(t)$ on $[a,b]$ satisfies $|\eta(t) - \eta(s)| \le L|t-s|$, $t, s \in [a,b]$, for some constant $L > 0$.*

(a) **Case with infinite n_i:** *Under* (2.3), *$n^{1/2}\left(\overline{AUC}(t) - \int_a^t \eta(s)\,ds\right)$ converges weakly to a zero-mean Gaussian process $\mathcal{G}_1(t)$ on $[a,b]$, with covariance function at t and t' equal to $E\left(\int_a^t v(s)\,ds \int_a^{t'} v(s)\,ds\right)$, as $n \to \infty$ and $\sum_{1\le i\le n}(nn_i)^{-1/2} \to 0$. The covariance function of $\mathcal{G}_1(t)$ can be estimated consistently by*

$$\hat{\psi}(t,t') = n\sum_{i=1}^{n} w_i(t)w_i(t')\left(AUC_i(t) - \overline{AUC}(t)\right)\left(AUC_i(t') - \overline{AUC}(t')\right). \tag{2.4}$$

(b) **Case with finite n_i:** *Assume that $E\left\{\sup_{a\le t\le b} Y^4(t)\right\} < \infty$ and $\{n_i\}$ are i.i.d. Then $n^{1/2}\left(\overline{AUC}(t) - E\{AUC(t)\}\right)$ converges weakly to a zero-mean Gaussian process $\mathcal{G}_2(t)$ on $[a,b]$, with covariance function at t and t' equal to $E\{[AUC(t) - E(AUC(t))][AUC(t') - E(AUC(t'))]\}$, as $n \to \infty$. The covariance function of $\mathcal{G}_2(t)$ can be estimated consistently by using* (2.4).

When the number of measurements per subject goes to infinity as in part (a), the individual $AUC_i(t)$ is mainly composed of the population mean AUC, $\int_a^t \eta(s)\,ds$, of all individual curves and the AUC of the random effect curve $v_i(t)$.

In the case when the number of measurements per subject is bounded, the contribution to the variance of $\overline{AUC}(t)$ by the measurement error part $e_i(t)$ cannot be ignored. The term $E\{AUC(t)\}$ is the mean AUC of all individual curves obtained by connecting observation points consisting of the observed time and response pairs.

3. Area Under the Curve Tests

The aim of this study is to develop simple nonparametric tests for comparison of nonparametric regression functions of longitudinal data without relying on smoothing. We consider two groups of data generated by (1.1) on the time interval $[a, b]$ and let superscript $k = 1, 2$ index the groups. The censoring times are generated independently from the distributions $P(C^{(k)} \le t)$, $k = 1, 2$, respectively. The tests are considered for two cases, (1) fixed and random designs with a large number of design time points, and (2) fixed and random designs with finite number of design time points. A nonparametric test for the null hypothesis $H_0: \ \eta^{(1)}(t) = \eta^{(2)}(t)$, $a \le t \le b$, may be based on comparing the weighted averages of AUCs for the two sets of longitudinal data. Define the process

$$U(t) = (n^{(1)} + n^{(2)})^{1/2} \left(\overline{AUC^{(1)}}(t) - \overline{AUC^{(2)}}(t) \right). \tag{3.1}$$

Various test statistics may be based on $U(t)$. For example, one may take $U(b)$ as a test statistic, it compares the weighted averages of areas under the curves for the two groups at the end of study. This test may lose power when the two regression functions $\eta^{(k)}(t)$, $k = 1, 2$, cross over each other and lose information on those curves that are censored before the end of study, but it should have good power otherwise. Under H_0 and conditions given in Theorem 2 below, this test statistic has an asymptotic normal distribution, with its variance estimated consistently by

$$\begin{aligned} \hat{\sigma}^2(b) = &(n^{(1)} + n^{(2)}) \sum_{i=1}^{n^{(1)}} (w_i^{(1)}(b))^2 \left(AUC_i^{(1)}(b) - \overline{AUC^{(1)}}(b) \right)^2 \\ &+ (n^{(1)} + n^{(2)}) \sum_{i=1}^{n^{(2)}} (w_i^{(2)}(b))^2 \left(AUC_i^{(2)}(b) - \overline{AUC^{(2)}}(b) \right)^2. \end{aligned} \tag{3.2}$$

Thus the test based on $U(b)/\hat{\sigma}(b)$ is a simple normal test and requires little computing time. In the following, we focus on

$$T = \sup_{a \le t \le b} |U(t)|. \tag{3.3}$$

The test based on T looks to detect any departure from the null hypothesis, the L^2-test is an alternative with a similar aim. Unlike the test based on $U(b)$, the asymptotic null distribution of T or the L^2-test statistic is intractable. Here we apply a Monte Carlo simulation method proposed by Lin, Wei and Ying (1993) to generate the distribution of the test statistics. Let $Z_1^{(k)}, \ldots, Z_{n^{(k)}}^{(k)}$, $k = 1, 2$, be i.i.d. standard normal random variables. Define

$$U^*(t) = ((n^{(1)}+n^{(2)})/n^{(1)})^{\frac{1}{2}} \Big[(n^{(1)})^{\frac{1}{2}} \sum_{i=1}^{n^{(1)}} w_i^{(1)}(t) \left(AUC_i^{(1)}(t) - \overline{AUC^{(1)}}(t)\right) Z_i^{(1)}\Big]$$
$$-((n^{(1)} + n^{(2)})/n^{(2)})^{\frac{1}{2}} \Big[(n^{(2)})^{\frac{1}{2}} \sum_{i=1}^{n^{(2)}} w_i^{(2)}(t) \left(AUC_i^{(2)}(t) - \overline{AUC^{(2)}}(t)\right) Z_i^{(2)}\Big]. \quad (3.4)$$

For both random and fixed designs, the distribution of the random process $U(t)$ can be approximated by the conditional distribution of the random process $U^*(t)$ given the observed data sequence.

Theorem 2. *Assume that the conditions of Theorem* 1 *hold for both regression models and that* $n^{(k)}/(n^{(1)} + n^{(2)}) \to \lambda_k$ *with* $0 < \lambda_k < 1$, $k = 1, 2$.

(a) **Case with infinite** n_i**:** *Under* H_0*, the processes* $U(t)$ *and* $U^*(t)$ *converge weakly to the same zero-mean Gaussian process on* $[a, b]$*, assuming that the conditions in* (a) *of Theorem* 1 *are satisfied for both data sets.*
(b) **Case with finite** n_i**:** *In addition to the conditions in* (b) *of Theorem* 1*, assume that* $n_i^{(1)}$ *and* $n_i^{(2)}$ *have the same distribution and that* $f^{(1)}(t) = f^{(2)}(t)$*. Then the processes* $U(t)$ *and* $U^*(t)$ *converge weakly to the same zero-mean Gaussian process on* $[a, b]$ *under* H_0*.*
In both (a) *and* (b)*, the weak convergence of* $U^*(t)$ *is regarded as the conditional convergence given the observed longitudinal data sequences.*

Let $T^* = \sup_{a \le t \le b} |U^*(t)|$. The critical value of the test statistic T of size α may be estimated by, say t_α, the $(1-\alpha)$ quantile of $\{T_r^*, r = 1, \ldots, B\}$ obtained by repeatedly generating B independent samples of i.i.d. normal random variables $Z_1^{(k)}, \ldots, Z_{n^{(k)}}^{(k)}$, $k = 1, 2$, while holding the observed data fixed, where T_r^* is the rth copy of T^*, $r = 1, \ldots, B$, and B is the number of repeated bootstrap samples. The null hypothesis H_0 is rejected at the significant level α if $T > t_\alpha$.

Consider $\Delta AUC(t) = \int_a^t (\eta^{(1)}(s) - \eta^{(2)}(s))\, ds$ under the conditions of (a) of Theorem 1, and $\Delta AUC(t) = E\{AUC^{(1)}(t)\} - E\{AUC^{(2)}(t)\}$ under the conditions of (b) of Theorem 1. From Theorem 1 and the proof of Theorem 2, $(n^{(1)} + n^{(2)})^{1/2}(\overline{AUC^{(1)}}(t) - \overline{AUC^{(2)}}(t) - \Delta AUC(t))$ and $U^*(t)$, given the observed data sequence, converge weakly to the same zero-mean Gaussian process. Hence, the maximum deviation test T is an omnibus test which is consistent against all alternatives H_a for which $\eta^{(1)}(t) \neq \eta^{(2)}(t)$ for some $t \in [a, b]$, provided the conditions in Theorem 1 are satisfied for both models.

Let $\hat{\sigma}^2(t)$ be similarly defined according to (3.2). From its uniform consistency under Theorem 1, an asymptotic $(1-\alpha)$ level uniform confidence band for $\Delta AUC(t)$ is given by

$$\left\{\overline{AUC^{(1)}}(t) - \overline{AUC^{(2)}}(t)\right\} \pm c_\alpha \hat{\sigma}(t)(n^{(1)} + n^{(2)})^{-1/2}, \tag{3.5}$$

where c_α is the estimate of the $(1-\alpha)$ quantile of $\sup_{a\le t\le b} |U^*(t)/\hat{\sigma}(t)|$. This last term can be obtained similarly by repeatedly generating i.i.d. normal random variables while holding the observed data fixed.

The procedures proposed above can be easily adapted to the one-sample case. The uniform confidence band for $\int_a^t \eta(s)\,ds$ or $E\{AUC(t)\}$ based on one sample longitudinal data is

$$\overline{AUC}(t) \pm c_\alpha (\hat{\psi}(t,t))^{1/2} n^{-1/2}, \tag{3.6}$$

where c_α is the estimate of the $(1-\alpha)$ quantile of $\sup_{a\le t\le b} |U_1^*(t)/(\hat{\psi}(t,t))^{1/2}|$, with $U_1^*(t) = n^{1/2}\sum_{i=1}^n w_i(t)\left(AUC_i(t) - \overline{AUC}(t)\right) Z_i$ and $\hat{\psi}(t,t)$ given in (2.4). Again, c_α can be estimated by repeatedly generating independent samples of i.i.d. normal random variables while holding the observed data fixed.

4. Simulations and An Example

In this section we present an extensive simulation study examining the levels and powers of the maximum deviation AUC test and apply the test to an AIDS clinical study comparing CD4+ counts under two different treatment regimens. Simulations are conducted for both fixed and random designs and for both uncensored and censored follow-up times.

4.1. Simulations

Let $Y^{(k)}(t) = \eta^{(k)}(t) + \epsilon^{(k)}(t)$, $0 \le t \le 1$, $k = 1, 2$, be the regression models from which two samples of longitudinal data are drawn, respectively. To examine the levels of the tests, the mean regression functions are taken to be $\sin(2\pi t)$ and $\sin(4\pi t)$. Four pairs of mean regression functions (see Table 2) are used to examine the powers. These alternatives attach different smoothness to the curves and different monotonicity. For convenience, we choose equal sample sizes for both groups. The number of subjects is taken to be $30, 50$ and the number of repeated measurements per subject is taken to be $5, 10, 20$. In Tables $1-3$, the first number under sample size is the number of subjects and the second number is the number of measurements per subject. For fixed design, each subject is followed up at fixed equal time intervals on $[0, 1]$ while for random design, the follow-up times for each subject are the ordered values of independent identically distributed uniform random variables on $[0, 1]$. For fixed design, censoring times are generated from a uniform $(0, 5)$ distribution, about 20% subjects drop out before the end of follow-up time ($t = 1$). For random uniform design, censoring times are generated from a uniform $(0, 0.5(1 - 1/(n_i + 1))$, distribution and this yields approximate 20% censoring, where n_i is the number of measurements for the ith subject.

Table 1. Empirical size for 500 simulations at 0.05 nominal level for fixed and random design, with error processes (1) and (2).

		0% censoring		20% censoring	
Function	Sample size	Error (1)	Error (2)	Error (1)	Error (2)
Fixed design					
$\sin(2\pi t)$	$(30, 5)$	6.4	5.8	8.0	5.4
	$(30, 10)$	5.4	5.2	6.0	5.6
	$(30, 20)$	4.4	3.8	5.2	3.6
	$(50, 5)$	5.4	5.4	5.6	4.2
	$(50, 10)$	5.2	4.8	5.0	5.8
	$(50, 20)$	5.8	6.8	5.6	6.6
$\sin(4\pi t)$	$(30, 5)$	4.2	6.2	5.4	5.6
	$(30, 10)$	6.0	6.2	5.2	4.2
	$(30, 20)$	4.8	3.2	8.2	6.2
	$(50, 5)$	6.2	4.8	6.0	5.0
	$(50, 10)$	5.2	4.4	4.4	7.8
	$(50, 20)$	6.8	6.0	5.0	5.4
Random design					
$\sin(2\pi t)$	$(30, 5)$	5.8	6.0	7.8	5.6
	$(30, 10)$	6.0	8.0	5.4	5.8
	$(30, 20)$	5.4	5.0	5.8	7.0
	$(50, 5)$	4.6	4.8	7.4	7.2
	$(50, 10)$	4.8	4.8	6.6	5.4
	$(50, 20)$	6.2	5.6	5.4	4.8
$\sin(4\pi t)$	$(30, 5)$	5.6	3.8	7.0	5.8
	$(30, 10)$	5.6	7.0	5.0	5.4
	$(30, 20)$	4.6	5.2	5.2	5.6
	$(50, 5)$	5.0	4.8	5.6	5.6
	$(50, 10)$	4.6	5.0	6.0	3.8
	$(50, 20)$	6.0	6.0	4.6	4.6

Let t_{ij} be the jth observation time of subject i in either of the two groups. Let $\{v_i\}$ be i.i.d. $N(0, \sigma_v^2)$ and $\{Z_{ij}\}$ be i.i.d. $N(0, \sigma_\epsilon^2)$, with $\sigma_v = 0.2$ and $\sigma_\epsilon = 0.5$. The following error processes are used in the simulation study:

Error process (1): $\epsilon^{(k)}(t_{ij}) = v_i + Z_{ij}$ for $k = 1, 2$. These error processes induce the same within-subject correlation and time independent variance-covariance

for both groups.
Error process (2): $\epsilon^{(k)}(t_{ij}) = v_i\eta^{(k)}(t_{ij}) + Z_{ij}$ for $k = 1, 2$. These error processes induce different within-subject correlation and time dependent variance-covariance for the two groups.

Table 2. Fixed design: empirical power for 500 simulations at 0.05 nominal level for fixed design, with error distributions (1) and (2).

		0% censoring		20% censoring	
Functions	Sample size	Error (1)	Error (2)	Error (1)	Error (2)
$\eta^{(1)}(t) = \sin(t)$					
$\eta^{(2)}(t) = \sqrt{t}$	$(30, 5)$	78.4	88.0	74.4	86.4
	$(30, 10)$	90.8	99.6	82.2	95.4
	$(30, 20)$	93.0	100	89.4	99.4
	$(50, 5)$	93.6	99.4	88.8	94.0
	$(50, 10)$	97.8	100	95.4	99.6
	$(50, 20)$	98.8	100	98.4	100
$\eta^{(1)}(t) = 0.3\sin(2\pi t)$					
$\eta^{(2)}(t) = 0.4\cos(3\pi t)$	$(30, 5)$	68.4	89.8	54.6	82.6
	$(30, 10)$	67.8	98.0	58.0	96.8
	$(30, 20)$	78.6	100	67.4	100
	$(50, 5)$	92.6	99.4	83.4	98.2
	$(50, 10)$	94.2	100	86.6	100
	$(50, 20)$	98.0	100	94.6	100
$\eta^{(1)}(t) = \exp(-t)$					
$\eta^{(2)}(t) = \exp(-2t)$	$(30, 5)$	76.2	88.0	65.4	77.6
	$(30, 10)$	82.6	97.6	77.8	93.4
	$(30, 20)$	93.0	100	88.8	99.4
	$(50, 5)$	92.0	98.2	83.8	92.0
	$(50, 10)$	97.0	100	92.8	99.8
	$(50, 20)$	98.8	100	96.8	100
$\eta^{(1)}(t) = \exp(-t)$					
$\eta^{(2)}(t) = (0.1\sqrt{t} + 1)$	$(30, 5)$	66.2	79.8	52.8	68.4
$\times \exp(-2t)$	$(30, 10)$	74.0	93.2	69.0	88.0
	$(30, 20)$	85.0	99.6	77.8	97.2
	$(50, 5)$	83.0	94.6	76.6	87.0
	$(50, 10)$	93.0	99.4	88.8	97.8
	$(50, 20)$	96.6	100	92.6	99.6

Table 1 contains the empirical sizes of the maximum deviation test at the 0.05 nominal level. Each entry in Table 1–3 is calculated based on 500 replicates and 500 bootstrap samples. Table 1 shows that the empirical sizes are reasonably close to the 0.05 nominal level for both fixed and random designs, presence or lack of censorship, two types of error process, and for different sample sizes.

Table 2 and Table 3 present the empirical powers of the proposed test for four different types of alternatives with various combinations of levels of censorship, error processes and sample sizes under fixed and random designs, respectively. The tables show that the AUC maximum deviation test has good power against different alternatives under both fixed and random designs. The powers under random design are consistently less than those under fixed designs, which shows a benefit of fixed design. Also, the power under the error process (2) is higher than under error process (1) since the variance of the error process (2) is smaller. Both error processes induce positive correlations between the responses at different times.

Table 3. Random design: empirical power for 500 simulations at 0.05 nominal level for random design, with error processes (1) and (2).

		0% censoring		20% censoring	
Functions	Sample size	Error (1)	Error (2)	Error (1)	Error (2)
$\eta^{(1)}(t) = \sin(t)$					
$\eta^{(2)}(t) = \sqrt{t}$	$(30, 5)$	66.6	81.4	57.8	72.8
	$(30, 10)$	81.2	95.6	73.4	91.2
	$(30, 20)$	93.4	99.4	82.0	98.2
	$(50, 5)$	87.4	96.8	73.0	85.4
	$(50, 10)$	97.6	99.8	92.6	99.4
	$(50, 20)$	98.8	100	96.8	100
$\eta^{(1)}(t) = 0.3\sin(2\pi t)$					
$\eta^{(2)}(t) = 0.4\cos(3\pi t)$	$(30, 5)$	12.8	15.2	10.4	14.2
	$(30, 10)$	32.2	72.6	21.4	62.0
	$(30, 20)$	63.4	99.8	48.8	98.6
	$(50, 5)$	14.0	23.8	14.8	23.6
	$(50, 10)$	57.8	96.8	44.8	88.6
	$(50, 20)$	94.8	100	84.0	100
$\eta^{(1)}(t) = \exp(-t)$					
$\eta^{(2)}(t) = \exp(-2t)$	$(30, 5)$	66.2	81.0	56.6	67.4
	$(30, 10)$	79.8	94.6	68.0	86.8
	$(30, 20)$	87.6	99.0	82.6	97.6
	$(50, 5)$	85.4	95.4	78.4	88.0
	$(50, 10)$	94.0	99.8	88.4	98.2
	$(50, 20)$	99.2	100	95.8	100
$\eta^{(1)}(t) = \exp(-t)$					
$\eta^{(2)}(t) = (0.1\sqrt{t} + 1)$	$(30, 5)$	56.2	70.4	45.6	58.6
$\times \exp(-2t)$	$(30, 10)$	72.6	86.2	61.0	77.0
	$(30, 20)$	77.4	96.2	74.0	93.8
	$(50, 5)$	77.8	88.6	68.2	79.8
	$(50, 10)$	90.0	97.0	82.0	95.0
	$(50, 20)$	96.4	100	89.4	99.6

4.2. Example

In an AIDS clinical study, the change in CD4+ cell counts is commonly used to assess the immunologic responses of anti-HIV treatment. We apply the AUC tests to the example introduced in Section 1, and compare the CD4+ responses (changes from baseline) between the two treatment groups (cohorts) of HIV-1 infected children. Seventeen patients in Cohort 1 were treated with a lower dose antiviral regimen and thirty-one patients in Cohort 2 were treated with the higher dose of the same regimen. First we consider 48 weeks treatment period. The proposed two-sample test gives a p-value of 0.0415, which shows a significant difference between the two treatment groups (in order to produce reliable results, two thousand bootstrap samples are used to calculate the p-values in this example). The pointwise mean response curves are plotted in Figure 1(b). We can see that Cohort 2 (higher dose treatment) performed better (more CD4+

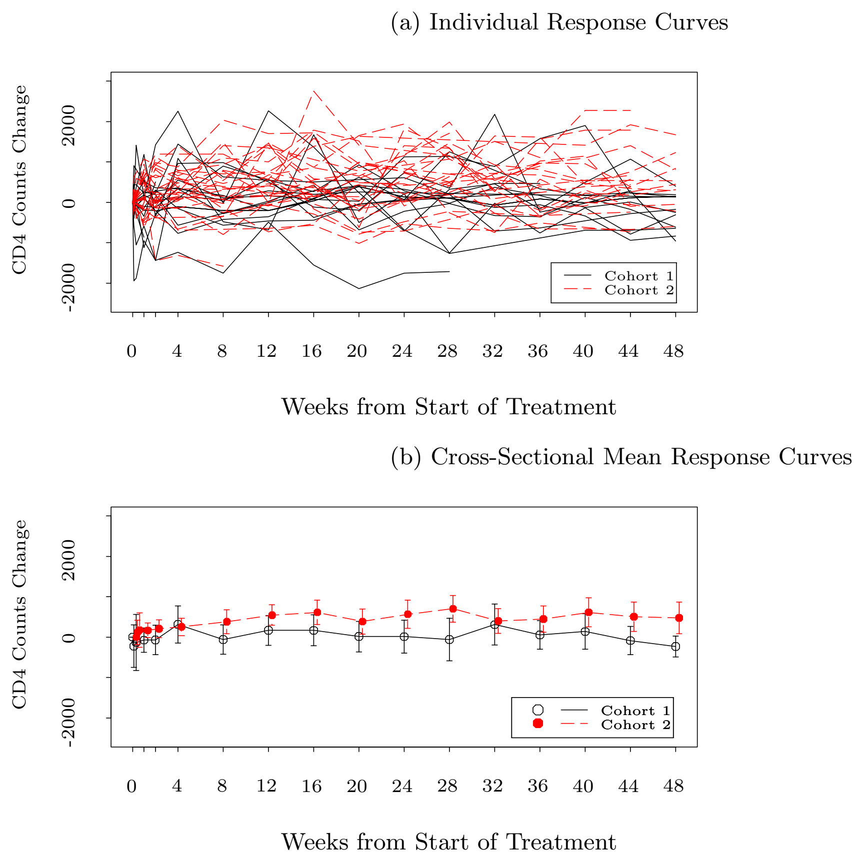

Figure 1. The change of CD4+ cell counts from baseline for the two treatment groups (cohorts) of HIV-1 infected children. The solid lines are the curves from cohort 1 (the low dose treatment) and the dashed lines are the curves from cohort 2 (the high dose treatment).

cells were recovered). The one-sample test is also conducted to see whether the response is significantly different from baseline. This shows that the response in Cohort 1 (lower dose treatment) is not significant (p-value= 0.267) while the response is significant (p-value< 0.0001) in Cohort 2 (higher dose treatment). When we repeat the analysis for the 24 weeks treatment, the two-sample test gives a p-value of 0.0865: they are marginally different. The p-values of the one-sample test are 0.0915 for Cohort 1 and 0.1025 for Cohort 2. There appears to be a benefit in CD4+ cell response from the higher dose of the study regimen for the 48 weeks treatment period, but the benefit has yet to become clear by the end of 24 weeks.

5. Appendix: Proofs

Let $U_1, \ldots, U_k$ be the order statistics of a random sample of size k from a uniform $(0,1)$ distribution. Define uniform spacings $\delta_i = U_i - U_{i-1}$, for $i = 1, \ldots, k+1$, where $U_0 \equiv 0$ and $U_{k+1} \equiv 1$. Let $\alpha_1, \ldots, \alpha_{k+1}$ be i.i.d. with an exponential distribution of mean one. Then $(\delta_1, \ldots, \delta_{k+1})$ and $(\alpha_1, \ldots, \alpha_{k+1})/\sum_{i=1}^{k+1} \alpha_i$ are equal in distribution (cf., Proposition 8.2.1, Shorack and Wellner (1986)). By this device one has a result on uniform spacings to be used in the proofs of Theorem 1 and 2. The proof is straight forward and is omitted.

Lemma 1.

$$E\left\{\max_{1\le i\le k+1} \delta_i\right\} = O\left(\log(k+1)/(k+1)\right)$$
$$E\left\{(\max_{1\le i\le k+1} \delta_i)^2\right\} = O\left((\log(k+1))^2/(k+1)^2\right).$$

Proof of Theorem 1. Part (a).

$$n^{1/2}\left(\overline{AUC}(t) - \int_a^t \eta(s)\,ds\right) = n^{1/2}\sum_{i=1}^n w_i(t)\left(AUC_i(t) - \int_a^t \eta(s)\,ds\right)$$
$$= n^{1/2}\sum_{i=1}^n w_i(t)\Big(AUC_i(t) - E[AUC_i(t)|\{t_{ij}\}]\Big)$$
$$+n^{1/2}\sum_{i=1}^n w_i(t)\left(E[AUC_i(t)|\{t_{ij}\}] - \int_a^t \eta(s)\,ds\right). \tag{5.1}$$

First we show that the second term on the right side of (5.1) converges to zero in probability uniformly in $t \in [a,b]$. Since $E[AUC_i(t)|\{t_{ij}\}] = \sum_{j=1}^{N_i(t)} [0.5(\eta(t_{ij-1}) + \eta(t_{ij}))(t_{ij} - t_{ij-1})] + \eta(t_{iN_i(t)})(t - t_{iN_i(t)})$, we have, by the uniform continuity of $\eta(t)$,

$$\sup_{a\le t\le b}\left|n^{1/2}\sum_{i=1}^n w_i(t)\left(E[AUC_i(t)|\{t_{ij}\}] - \int_a^t \eta(s)\,ds\right)\right|$$

$$\leq O(1) \sup_{a\leq t\leq b} \sum_{i=1}^{n} n^{1/2} w_i(t) \max_{1\leq j\leq n_i} (t_{ij} - t_{ij-1})$$
$$\leq O_p(1) n^{-1/2} \sum_{i=1}^{n} \max_{1\leq j\leq n_i} (t_{ij} - t_{ij-1}). \tag{5.2}$$

Let $F(t)$ be the distribution function associated with the density $f(t)$ of $\{t_{ij}^0\}$. Let $U_{ij} = F(t_{ij})$, $1 \leq i \leq n$, $1 \leq j \leq n_i$. Then $\{U_{ij} - U_{ij-1}, j = 1, \ldots, n_i + 1\}$ are uniform spacings with $U_{i,n_i+1} \equiv 1$. The sets of uniform spacings are independent among different subjects. By the continuity of $f(t)$, $U_{ij} - U_{ij-1} = f(t_{ij}^*)(t_{ij} - t_{ij-1})$, where t_{ij}^* is on the line segment between t_{ij-1} and t_{ij}. Thus, $t_{ij} - t_{ij-1} \leq O(1)(U_{ij} - U_{ij-1})$ since $f(t)$ is bounded away from zero. From (5.2), it follows

$$\sup_{a\leq t\leq b} \Big| n^{1/2} \sum_{i=1}^{n} w_i(t) \left(E[AUC_i(t)|\{t_{ij}\}] - \int_a^t \eta(s)\, ds \right) \Big|$$
$$\leq O_p(1) n^{-1/2} \sum_{i=1}^{n} \max_{1\leq j\leq n_i} (U_{ij} - U_{ij-1}), \tag{5.3}$$

which converges to zero in probability by applying the Markov inequality and Lemma 1.

Now, we prove the weak convergence of the first term in (5.1). Let $X_i(t) = AUC_i(t)–E\{AUC_i(t)|\{t_{ij}\}\}$ and, for a function $h(t)$, let $I_h(t) = \sum_{j=1}^{N_i(t)} \{0.5[h(t_{ij-1}) + h(t_{ij})](t_{ij} - t_{ij-1})\} + h(t_{iN_i(t)})(t - t_{iN_i(t)})$. We have

$$AUC_i(t) - E[AUC_i(t)|\{t_{ij}\}] = I_{\epsilon_i}(t) = I_{v_i}(t) + I_{e_i}(t). \tag{5.4}$$

Note that $|I_{v_i}(t) - \int_a^t v_i(s)\, ds| \leq L_i \max_{1\leq j\leq n_i} |t_{ij} - t_{ij-1}|$ and so

$$\sup_{a\leq t\leq b} \Big| n^{1/2} \sum_{i=1}^{n} w_i(t) \Big(I_{v_i}(t) - \int_a^t v_i(s)\, ds \Big) \Big|$$
$$\leq \sup_{a\leq t\leq b} \Big| n^{1/2} \sum_{i=1}^{n} w_i(t) L_i \max_{1\leq j\leq n_i} (t_{ij} - t_{ij-1}) \Big|$$
$$\leq O_p(1) n^{-1/2} \sum_{i=1}^{n} L_i \max_{1\leq j\leq n_i} (U_{ij} - U_{ij-1}) \to 0, \tag{5.5}$$

in probability by the Markov inequality and the Cauchy-Schwarz inequality.

From Scheike (2000, cf., the proof of Proposition 1), it follows that ${n_i}^{1/2} I_{e_i}(t)$ converges weakly to a mean zero Gaussian martingale as $n_i \to \infty$. Thus

$$\sup_{a\leq t\leq b} \Big| n^{1/2} \sum_{i=1}^{n} w_i(t) I_{e_i}(t) \Big| = \sup_{a\leq t\leq b} \Big| n^{1/2} \sum_{i=1}^{n} n_i^{-1/2} w_i(t) {n_i}^{1/2} I_{e_i}(t) \Big|$$

$$\le O_p(1)n^{-1}\sum_{i=1}^{n}(n/n_i)^{1/2}\sup_{a\le t\le b}|{n_i}^{1/2}I_{e_i}(t)|$$
$$\to 0 \quad \text{in probability as } n^{-1}\sum_{i=1}^{n}(n/n_i)^{1/2}\to 0, \tag{5.6}$$

since the distributions of $\sup_{a\le t\le b}|{n_i}^{1/2}I_{e_i}(t)|$ are the same for different individuals with the same n_i. From (5.4) to (5.6), it follows

$$n^{1/2}\sum_{i=1}^{n} w_i(t)\Big(AUC_i(t)-E[AUC_i(t)|\{t_{ij}\}]\Big)$$
$$=\left(\sum_{i=1}^{n}\xi_i(t)\right)^{-1} n^{1/2}\sum_{i=1}^{n}\xi_i(t)\int_a^t v_i(s)\,ds+o_p(1),$$

which converges weakly to a mean zero Gaussian process on $[a,b]$ with covariance function at t and t' equal to $E(\int_a^t v_i(s)ds\int_a^{t'} v_i(s)ds)$. The weak convergence follows from Example 2.11.14 of van der Vaart and Wellner (1996).

Finally, since

$$n^{-1}\sum_{i=1}^{n}\Big(\xi_i(t)\int_a^t v_i(s)ds\xi_i(t')\int_a^{t'} v_i(s)ds\Big)\to E\Big(\xi_i(t)\int_a^t v_i(s)ds\xi_i(t')\int_a^{t'} v_i(s)ds\Big),$$

we have, by (5.4) to (5.6),

$$n^{-1}\sum_{i=1}^{n}\Big(\xi_i(t)(AUC_i(t)-E[AUC_i(t)|\{t_{ij}\}])\xi_i(t')(AUC_i(t')-E[AUC_i(t')|\{t_{ij}\}])\Big)$$
$$\to E\Big(\xi_i(t)\int_a^t v_i(s)\,ds\;\xi_i(t')\int_a^{t'} v_i(s)\,ds\Big).$$

Since $\sup_{a\le t\le b} n^{-1}\sum_{i=1}^{n}|E[AUC_i(t)|\{t_{ij}\}]-\int_a^t\eta(s)ds|\le O(1)n^{-1}\sum_{i=1}^{n}\max_{1\le j\le n_i}(t_{ij}-t_{ij-1})\to 0$ in probability and $\overline{AUC}(t)-\int_a^t\eta(s)\,ds$ converges to zero in probability uniformly in $t\in[a,b]$ from the previous weak convergence result, we have

$$n^{-1}\sum_{i=1}^{n}\left(\xi_i(t)\left(AUC_i(t)-\overline{AUC}(t)\right)\,\xi_i(t')\left(AUC_i(t')-\overline{AUC}(t')\right)\right)$$
$$\to E\Big(\xi_i(t)\int_a^t v_i(s)\,ds\;\xi_i(t')\int_a^{t'} v_i(s)\,ds\Big).$$

Thus, by $n^{-1}\sum_{i=1}^{n}\xi_i(t)\to E\xi_i(t)$ and the independence between $\xi_i(t)$ and $Y_i(t)$, we have $\hat\psi(t,t')\to E(\int_a^t v_i(s)ds\int_a^{t'} v_i(s)ds)$. This completes the proof of part (a).

Part (b). Start with

$$n^{1/2}\left(\overline{AUC}(t)-E[AUC(t)]\right)$$
$$=n\Big(\sum_{i=1}^{n}\xi_i(t)\Big)^{-1}n^{-1/2}\sum_{i=1}^{n}\xi_i(t)\Big(AUC_i(t)-E[AUC(t)]\Big).$$

Let $X_i(t)=AUC_i(t)-E[AUC(t)]$. Since $\sum_{i=1}^n \xi_i(t)/n$ converges uniformly to $P(C_i\ge t)$ which is bounded away from zero, it suffices to show the weak convergence of $n^{-1/2}\sum_{i=1}^n \xi_i(t)\ \times X_i(t)$. Note that $\{\xi_i(t)X_i(t)\}$ are i.i.d. random processes. Since $E\left\{\sup_{a\le t\le b}Y^2(t)\right\}<\infty$, we have that $\sup_{1\le i\le n}|E[X_i(t)|\{t_{ij}\}]|$ is bounded and $E|\xi_i(t)X_i(t)-\xi_i(s)X_i(s)|\le E\{\xi_i(t)E(|X_i(t)-X_i(s)|\ |\{t_{ij}\})\}+E\{|\xi_i(t)-\xi_i(s)|\ \ E(|X_i(s)|)|\{t_{ij}\}\}$ is bounded by $K(t-s)$ for some constant K. The weak convergence of $n^{1/2}(\overline{AUC}(t)-\ E[AUC(t)])$ to a Gaussian process follows from Example 2.11.14 of van der Vaart and Wellner (1996) and the covariance function at t and t' is equal to $E\{(AUC(t)-E[AUC(t)])\ (AUC(t')-E[AUC(t')])\}$.

By the law of large numbers, we have $\overline{AUC}(t)\to E\{AUC(t)\}$ and $\hat{\psi}(t,t')\to E\{(AUC(t)\ -E[AUC(t)])\ (AUC(t')-E[AUC(t')])\}$ in probability as $n\to\infty$. This completes the proof of part (b).

Proof of Theorem 2.

First note that the two groups of longitudinal data sets are independent. Under H_0, in view of the results from Theorem 1, it is sufficient to show $U_1^*(t)=n^{1/2}\sum_{i=1}^n w_i(t)(AUC_i(t)-\overline{AUC}(t))Z_i$ converges weakly to $\mathcal{G}_1(t)$ or $\mathcal{G}_2(t)$ under the conditions (a) or (b) of Theorem 1, respectively, given the observed data sequence, where $Z_1,\ldots,Z_n$ are i.i.d. standard normal random variables. Further, in view of the results of Theorem 1 on the consistent estimate of the covariance functions, and by the Multivariate Central Limit Theorem, the conditional finite-dimensional distributions of $U_1^*(t)$ converge to the finite-dimensional distributions of $\mathcal{G}_1(t)$ or $\mathcal{G}_2(t)$ under the conditions of (a) or (b) of Theorem 1, respectively. Thus, it is left to prove that $U_1^*(t)$ is tight under either of the conditions (a) and (b) of Theorem 1.

Let $\zeta_i(t)=\xi_i(t)\left(AUC_i(t)-\overline{AUC}(t)\right)$. Considering the fact of uniform convergence of $\sum_{i=1}^n\xi_i(t)/n$, it suffices to show that the process $n^{-1/2}\sum_{i=1}^n\zeta_i(t)Z_i$ is tight given the observed data sequence. We check a slight extension of the moment conditions of Theorem 15.6 of Billingsley (1968) for the tightness by McKeague and Zhang (1994, p.507). Here we state the corrected version of the relaxed moment condition for the process $X_n(t)$ to be tight, obtained through conversations with the authors:

$$E\{|X_n(t)-X_n(t_1)|^\gamma|X_n(t_2)-X_n(t)|^\gamma\}\le(F(t_2)-F(t_1))^{2\alpha}+o(1)(F(t_2)-F(t_1)),\tag{5.7}$$

where $t_1 \le t \le t_2$, $\gamma > 0$, $\alpha > 1/2$, $o(1)$ converges to zero uniformly in (t_1, t, t_2) and F is a nondecreasing and continuous function. Applying (5.7), conditional on the observed data sequence, we have, for $a \le t_1 \le t \le t_2 \le b$,

$$\begin{aligned}
& n^{-2} E\Big\{\Big(\sum_{i=1}^n (\zeta_i(t) - \zeta_i(t_1)) Z_i\Big)^2 \Big(\sum_{i=1}^n (\zeta_i(t_2) - \zeta_i(t)) Z_i\Big)^2 \Big| \{\text{observed data}\}\Big\} \\
\le\ & 2n^{-2} \sum_{i=1}^n (\zeta_i(t) - \zeta_i(t_1))^2 \sum_{i=1}^n (\zeta_i(t_2) - \zeta_i(t))^2 \\
& + 4n^{-2} \Big(\sum_{i=1}^n (\zeta_i(t) - \zeta_i(t_1))(\zeta_i(t_2) - \zeta_i(t))\Big)^2 \\
& + 3n^{-2} \sum_{i=1}^n (\zeta_i(t) - \zeta_i(t_1))^2 (\zeta_i(t_2) - \zeta_i(t))^2 \\
\le\ & 6n^{-2} \sum_{i=1}^n (\zeta_i(t) - \zeta_i(t_1))^2 \sum_{i=1}^n (\zeta_i(t_2) - \zeta_i(t))^2 \\
& + 3n^{-2} \sum_{i=1}^n (\zeta_i(t) - \zeta_i(t_1))^2 (\zeta_i(t_2) - \zeta_i(t))^2.
\end{aligned}$$

Under the conditions of part (a), Theorem 1, we have

$$\begin{aligned}
& n^{-2} \sum_{i=1}^n (\zeta_i(t) - \zeta_i(t_1))^2 \sum_{i=1}^n (\zeta_i(t_2) - \zeta_i(t))^2 \\
\to\ & E\Big(\xi_i(t) \int_a^t v_i(s)\, ds - \xi_i(t_1) \int_a^{t_1} v_i(s)\, ds\Big)^2 \\
& \times E\Big(\xi_i(t_2) \int_a^{t_2} v_i(s)\, ds - \xi_i(t) \int_a^{t} v_i(s)\, ds\Big)^2 \\
\le\ & 4\Big[\int_{t_1}^t dF_c(s) E\Big(\int_a^t v_i(s)\, ds\Big)^2 + E\Big(\int_{t_1}^t v_i(s)\, ds\Big)^2\Big] \\
& \times \Big[\int_t^{t_2} dF_c(s) E\Big(\int_a^{t_2} v_i(s)\, ds\Big)^2 + E\Big(\int_t^{t_2} v_i(s)\, ds\Big)^2\Big] \\
\le\ & K\Big(\int_{t_1}^{t_2} dF_c(s) + \int_{t_1}^{t_2} E(v_i^2(s))\, ds\Big)^2, \qquad (5.8)
\end{aligned}$$

for some constant K uniformly in (t_1, t, t_2), where $F_c(t)$ is the distribution function of the censoring time C. Similarly,

$$\begin{aligned}
& n^{-1} \sum_{i=1}^n (\zeta_i(t) - \zeta_i(t_1))^2 (\zeta_i(t_2) - \zeta_i(t))^2 \\
\to\ & E\Big[\Big(\xi_i(t) \int_a^t v_i(s) ds - \xi_i(t_1) \int_a^{t_1} v_i(s) ds\Big)^2 \Big(\xi_i(t_2) \int_a^{t_2} v_i(s) ds - \xi_i(t) \int_a^t v_i(s) ds\Big)^2\Big]
\end{aligned}$$

$$\leq K'\Big(\int_{t_1}^{t_2} dF_c(s) + \int_{t_1}^{t_2} E(v_i^4(s))\, ds\Big), \tag{5.9}$$

for some constant K' uniformly in (t_1, t, t_2). Then (5.8) and (5.9) imply that the process $n^{-1/2}\sum_{i=1}^n \zeta_i(t)Z_i$ is tight given the observed data sequence.

Under the conditions of part (b), Theorem 1,

$$\begin{aligned} & n^{-2}\sum_{i=1}^n(\zeta_i(t)-\zeta_i(t_1))^2\sum_{i=1}^n(\zeta_i(t_2)-\zeta_i(t))^2 \\ & \to E\Big\{\xi_i(t)[AUC_i(t)-E(AUC_i(t))]-\xi_i(t_1)[AUC_i(t_1)-E(AUC_i(t_1))]\Big\}^2 \\ & \quad \times E\Big\{\xi_i(t_2)[AUC_i(t_2)-E(AUC_i(t_2))]-\xi_i(t)[AUC_i(t)-E(AUC_i(t))]\Big\}^2, \end{aligned}$$

uniformly in (t_1, t, t_2). The first term of the above product is bounded by $2E\{\xi_i(t)-\xi_i(t_1)\}^2 E\{AUC_i(t)-E[AUC_i(t)]\}^2 + 2E(\{AUC_i(t)-AUC_i(t_1)\} - E\{AUC_i(t)-AUC_i(t_1)\})^2 \leq K(P(t_1 \leq C < t) + (t-t_1))$, for some constant K, under the condition $E\left\{\sup_{a\leq t\leq b} Y^4(t)\right\} < \infty$. Similar result holds for the second term of the product. It also follows that

$$\begin{aligned} & n^{-1}\sum_{i=1}^n(\zeta_i(t)-\zeta_i(t_1))^2(\zeta_i(t_2)-\zeta_i(t))^2 \\ & \to E\Big\{\Big(\xi_i(t)[AUC_i(t)-E(AUC_i(t))]-\xi_i(t_1)[AUC_i(t_1)-E(AUC_i(t_1))]\Big)^2 \\ & \quad \times\Big(\xi_i(t_2)[AUC_i(t_2)-E(AUC_i(t_2))]-\xi_i(t)[AUC_i(t)-E(AUC_i(t))]\Big)^2\Big\} \\ & \leq K'\Big(P(t_1 \leq C < t_2) + (t_2 - t_1)\Big), \end{aligned}$$

for some constant K' uniformly in (t_1, t, t_2). Therefore, the process $U_1^*(t)$ is tight under either of the conditions (a) and (b) of Theorem 1, given the observed data sequence.

Acknowledgements

The research was partially supported by NIH grants AI43220, AI45356, AI52765 and AI38855 (Hulin Wu) and NSF grant DMS-0304922 (Yanqing Sun). The authors are grateful to Ms. Ping Ruan for her help in data preparation and plots. The authors also thank two referees for insightful comments and suggestions.

References

Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

Davidian, M. and Giltinan, D. (1995). *Nonlinear Models for Repeated Measurement Data.* Chapman and Hall, New York.

Diggle P. J., Liang K. Y. and Zeger S. L. (1994). *Analysis of Longitudinal Data.* Oxford University Press, New York.

Hoover, D. R., Rice, J. A., Wu, C. O. and Yang, L. P. (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika* **85**, 809-822.

Hoover, D. R. and Wu, C. O. (2001). Comment on "Semiparametric and nonparametric regression analysis of longitudinal data", by Lin, D. Y. and Ying, Z. *J. Amer. Statist. Assoc.* **96**, 116-119.

Lin, D. Y., Wei, L. J. and Ying, Z. (1993). Checking the Cox model with cumulative sums of martingale-based residuals. *Biometrika* **80**, 557-572.

Lin, D. Y. and Ying, Z. (2001). Semiparametric and nonparametric regression analysis of longitudinal data. *J. Amer. Statist. Assoc.* **96**, 103-113.

Martinussen, T. and Scheike, T. H. (1999). A semiparametric additive regression model for longitudinal data. *Biometrika* **86**, 691-702.

Martinussen, T. and Scheike, T. H. (2000). A nonparametric dynamic additive regression model for longitudinal data. *Ann. Statist.* **28**, 1000-1025.

Martinussen, T. and Scheike, T. H. (2001). Sampling adjusted analysis of dynamic additive regression models for longitudinal data. *Scand. J. Statist.* **28**, 303-323.

McKeague, I. W. and Zhang, M. J. (1994). Identification of nonlinear time series from first order cumulative characteristics. *Ann. Statist.* **22**, 495-514

Moyeed, R. A. and Diggle, P. J. (1994). Rates of convergence in semiparametric modelling of longitudinal data. *Austral. J. Statist.* **36**, 75-93.

Rice, J. A. and Wu, C. O. (2001). Nonparametric mixed effects models for unequally sampled noise curves. *Biometrics* **57**, 253-259.

Rowland, M. and Tozer, T. N. (1995). *Clinical Pharmacokinetics: Concepts and Applications.* 3rd edition. Williams and Wilkins, Baltimore.

Scheike, T. H. (2000). Comparison of non-parametric regression functions through their cumulatives. *Statist. Probab. Lett.* **46**, 21-32.

Scheike, T. H. and Zhang, M. J. (1998). Cumulative regression function tests for longitudinal data. *Ann. Statist.* **26**, 1328-1355.

Scheike, T. H., Zhang, M. J. and Juul, A. (1999). Comparing reference charts. *Biometrical J.* **41**, 679-687.

Shi, M., Weiss, R. E. and Taylor, J. M. G. (1996). An analysis of pediatric CD4 counts for acquired immune deficiency syndrome using flexible random curves. *Appl. Statist.* **45**, 151-163.

Shorak, G. R. and Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics.* John Wiley and Sons, Inc., New York.

Van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes with Applications to Statistics.* Springer, New York.

Vonesh, E. F. and Chinchilli, V. M. (1996). *Linear and Nonlinear Models for the Analysis of Repeated Measurements.* Marcel Dekker, New York.

Weinberg J. M. and Lagakos S. W. (2000). Efficiency comparisons of rank and permutation tests based on summary statistics computed from repeated measures data. *Statist. Medicine* **20**, 705-731.

Wu, C. O., Chiang, C. T. and Hoover, D. R. (1998). Asymptotic confidence regions for kernel smoothing of a time-varying coefficient model with longitudinal data. *J. Amer. Statist. Assoc.* **88**, 1388-1402.

Wu, H. and Zhang, J. T. (2002). Local polynomial mixed-effects models for longitudinal data. *J. Amer. Statist. Assoc.* **97**, 883-897.

Zeger, S. L. and Diggle, P. J. (1994). Semiparametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters. *Biometrics* **55**, 452-459.

Department of Mathematics, University of North Carolina, Charlotte, NC 28223, U.S.A.

E-mail: yasun@uncc.edu

Department of Biostatistics and Computational Biology, University of Rochester Medical Center, 601 Elmwood Avenue, Box 630 Rochester, NY 14642, U.S.A.

E-mail: Hulin_Wu@urmc.rochester.edu

(Received December 2001; accepted April 2003)