



A nonparametric test for the evaluation of group sequential clinical trials with covariate information



Ao Yuan^{a,*}, Yanxun Zheng^a, Peng Huang^b, Ming T. Tan^a

^a Department of Biostatistics, Bioinformatics and Biomathematics, Georgetown University, Washington, DC 20057, USA

^b Department of Oncology, Johns Hopkins University, Baltimore, MD 21218, USA

ARTICLE INFO

Article history:

Received 30 October 2015

Available online 22 August 2016

AMS subject classifications:

primary 62K05

secondary 62H99

Keywords:

Covariate

Mann–Whitney difference

Group sequential clinical trial

Sequential conditional probability ratio test

boundary

U-statistic

ABSTRACT

Group sequential design is frequently used in clinical trials to evaluate a new treatment vs a control. Although nonparametric methods have the advantage of robustness, most such methods do not take into consideration of covariate information that could be used to improve the test accuracy if incorporated properly. We address this problem using a two-sample U-statistic that incorporates covariate information into the test statistic. The asymptotic properties of the proposed estimator are presented. Simulations are conducted to evaluate the performance of the test. We then apply the proposed method to the analysis of data from a Parkinson disease clinical trial, and demonstrate that the significance of the effect associated with deprenyl could be detected at an early stage.

© 2016 Elsevier Inc. All rights reserved.

1. Introduction

Group sequential designs are commonly used in clinical trials to compare new treatments against existing ones; see, e.g., Jennison and Turnbull [7] or Kosorock et al. [12]. Multiple primary end points are increasingly used for such an evaluation. As parametric-based multivariate inference is often not robust to model specification, however, nonparametric methods should be favored in this regard. In addition, the frequently used assumption of identical variances of the responses under the null hypothesis is violated in many clinical trial settings. This is the so-called Behrens–Fisher problem for finite-sample-size tests.

For the analysis of non-sequential trial designs, Huang et al. [5] modified O'Brien's nonparametric rank test [16] to avoid the Behrens–Fisher problem. Huang et al. [6] further proposed a measure of global treatment effect (GTE) for identifying more beneficial effects when multiple primary endpoints are used; they also showed how to carry out sample size computation. Liu et al. [14] proposed a maximization test to identify treatment difference (either beneficial or detrimental). Lin, Li and Tan [13] proposed a rank regression method to estimate the covariate effects on combined survival and quantitative data. Huang and Tan [4] proposed a rank-based method for group sequential design with multiple primary endpoints where the GTE was used as the primary parameter of interest. There are many other methods for comparing two groups; see, e.g., Koch et al. [9,10] and Schacht et al. [21].

When covariate information is available, it could conceivably be used to improve the test accuracy. Throughout this paper, the term “covariate” refers to baseline covariates, i.e., taken to be independent of the treatment assignment. There is extensive literature on the issue of whether or not, and how to, incorporate covariate information, to adjust the analysis

* Corresponding author.

E-mail address: ay312@georgetown.edu (A. Yuan).

of treatment difference, and to increase the precision of the test statistic. Tsiatis et al. [25] studied the related issues and proposed a semiparametric approach which generalizes the traditional covariate adjustment method. Lu and Tsiatis [15] investigated this problem for survival data. Zhang et al. [30] investigated model-based parametric methods to incorporate covariates to enhance the estimation efficiency in the context of clinical trials.

A typical baseline covariate adjustment method in the regression setting is the following. Let $Y_1 = (y_{1,1}, \dots, y_{1,n_1})^\top$ and $Y_2 = (y_{2,1}, \dots, y_{2,n_2})^\top$ be the observed responses of two groups, and $X_1 = (x_{1,1}, \dots, x_{1,n_1})$ and $X_2 = (x_{2,1}, \dots, x_{2,n_2})$ be the corresponding covariates. Denote by $\bar{Y}_1, \bar{Y}_2, \bar{X}_1, \bar{X}_2$ the corresponding sample means, and set $\Omega = \Omega_1 + \Omega_2$, where, for $k \in \{1, 2\}$, Ω_k is the sample covariance matrix of $(Y_k^\top, X_k^\top)^\top$. Partition Ω as Ω_{yy}, Ω_{yx} and Ω_{xx} . Without covariates, the difference is $\bar{Y}_1 - \bar{Y}_2$, and its variance is Ω_{yy} . In the presence of covariates, the adjusted difference is $(\bar{Y}_1 - \bar{Y}_2) - \Omega_{xy}\Omega_{xx}^{-1}(\bar{X}_1 - \bar{X}_2)$; the variance of the adjusted difference is $\Omega_{yy} - \Omega_{xy}\Omega_{xx}^{-1}\Omega_{yx}$, which typically exhibits a reduction in variance due to the incorporation of the covariate information. As using the full data is often more efficient than using ranks, we are interested in using a U-type statistic to estimate the Mann–Whitney difference θ . However, for this particular problem, estimating θ from the adjusted data $(y_{1i} - \Omega_{xy}\Omega_{xx}^{-1}x_{1i}, y_{2i} - \Omega_{xy}\Omega_{xx}^{-1}x_{2i})$ with $i \in \{1, \dots, n\}$ is not appealing, as this would rather lead to estimate

$$\Pr(Y_1 - \Omega_{xy}\Omega_{xx}^{-1}X_1 < Y_2 - \Omega_{xy}\Omega_{xx}^{-1}X_2) - \Pr(Y_2 - \Omega_{xy}\Omega_{xx}^{-1}X_2 < Y_1 - \Omega_{xy}\Omega_{xx}^{-1}X_1),$$

not θ itself. Also, many of the covariates in clinical trial are population variables. Often, there is side information available and incorporating such information in the estimation of θ by empirical weighing is more natural and simpler than the above method.

Our objective here is to design a robust and easy method for treatment comparison with multiple primary endpoints in the multistage clinical trial setting, and to adjust potential confounding covariates without the assumption of identical variances. To this end, we propose a U-statistic based method to estimate the Mann–Whitney difference between two comparing treatments, and we use the sequential conditional probability ratio test boundary [23,28] to monitor the trial progress. We consider two cases when covariate auxiliary information is present: incorporating the information into the test statistic by empirical weighing, and dealing with the case of no auxiliary information in parallel. We show that the proposed method is efficient and easy to use. Simulation studies are conducted to evaluate the performance of the proposed method, and then the method is used to analyze a real clinical trial on Parkinson disease, where our method allows to detect the effects at an early stage with the proposed group sequential trial. Thus considerable human and financial costs can be saved with the proposed group sequential trial. All the relevant proofs are given in the [Appendix](#).

2. The problem

This work is motivated by an extension of the group sequential test from Huang and Tan [4]. In group sequential clinical trials, the goal is to evaluate, at each stage, whether a new treatment is more efficacious than some existing one based on the observed responses of patients from the two treatments. For all $j \in \{1, 2\}$, $t \in \mathbb{N}$, and $i \in \{1, \dots, n_{jt}\}$, let $y_{ji} \in \mathbb{R}^k$ be the observed response of the i th subject, in the j th treatment, at the t th stage (interim). Here for fixed j , the y_{ji} 's are assumed i.i.d. over i , and the observations are coded in such a way that larger values correspond to better treatment effects. In group sequential clinical trials, the number of stages is often between 2 and 5 although it can go up to 10. Here we make it general without imposing an upper bound for it.

The t th interim analysis will be performed when data from the first n_{jt} subjects are available. Thus the cumulative sample size n_{jt} at each stage t satisfies $n_{jt} < n_{j,t+1}$ for all (j, t) . The goal is to compare the effects of the two treatments based on the observed responses y_{ji} 's, to investigate whether one of the treatments is more efficacious than the other. For this, let $Y_1 = (Y_{11}, \dots, Y_{1k})^\top$ and $Y_2 = (Y_{21}, \dots, Y_{2k})^\top$ be i.i.d. copy of y_{1i} and y_{2i} . Commonly used parameters for this objective are the Mann–Whitney differences among the components of Y_1 and Y_2 , defined for all $j \in \{1, \dots, k\}$, by

$$\theta_j = \Pr(Y_{1j} < Y_{2j}) - \Pr(Y_{1j} > Y_{2j}).$$

Let $\mathbf{a} = (a_1, \dots, a_k)^\top$ be a given vector with positive components, and $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)^\top$. We formulate the problem as a statistical hypothesis test of

$$\mathcal{H}_0 : \bar{\theta} = \mathbf{a}^\top \boldsymbol{\theta} = \sum_{j=1}^k a_j \theta_j \leq 0 \quad \text{vs.} \quad \mathcal{H}_1 : \bar{\theta} > 0.$$

Setting $a_j = 1/k$ for all j , then $\bar{\theta}$ is the global treatment effect in Huang and Tan [4]. When there are no covariate observations, they proposed a rank statistic for testing \mathcal{H}_0 . Although we concentrate on testing the one-sided hypothesis \mathcal{H}_0 , the proposed method can actually handle two-sided or any other form of composite hypothesis.

Here we consider the same problem as above in the presence of covariate data. Let $x_{ji} \in \mathbb{R}^d$ be the observed covariates for y_{ji} . We propose a two-sample U-statistic to incorporate the covariates x_{ji} 's into the estimation of $\bar{\theta}$, as described below.

3. The proposed method and its properties

3.1. The method

Let $z_{ji} = (y_{ji}^\top, x_{ji}^\top)^\top$ be the combined response-covariate $(k+d)$ -dimensional observations. For $j \in \{1, 2\}$, the z_{ji} 's are i.i.d. across i with common unknown distribution function F_j . Let $\mathbf{1}(\cdot)$ be the indicator function, and

$$h(Y_1, Y_2) = \sum_{j=1}^k a_j \{\mathbf{1}(Y_{1j} < Y_{2j}) - \mathbf{1}(Y_{1j} > Y_{2j})\}.$$

Note that h is symmetric within the components of Y_1 and Y_2 , respectively. Therefore, it is a valid kernel for a two-sample U-statistic. Let $X = (X_1, \dots, X_d)^\top$ be an i.i.d. copy of the x_{ji} 's. Below we distinguish two cases, according to auxiliary information from the covariates is present or not.

As the covariates are population variables, investigated by numerous studies, often there is auxiliary (side) information available about some or all of them. This can be written in terms of a known function $g(X) = (g_1(X), \dots, g_r(X))^\top$ with r constraints, namely $E\{g(X)\} = \mathbf{0}$. If g contains a d -dimensional unknown parameter, we must have $r \geq d$. If we know the mean $\mu = (\mu_1, \dots, \mu_d)^\top$ of the covariates, then we can take $g(X) = (X_1 - \mu_1, \dots, X_d - \mu_d)^\top$.

To incorporate the auxiliary information into the proposed statistic, we will rely on the empirical likelihood (EL) method of Owen [18,19]. More specifically, we will combine the method for incorporating auxiliary information into the EL due to Qin and Lawless [20], and the method of Yuan et al. [29] for one-sample U-statistics with covariate information. Here we construct a two-sample U-statistic for the parameter θ using the kernel h and the auxiliary information constraints given by g . We need the following notations.

Let $\mathbf{i} = (i_1, i_2)^\top$, $\mathbf{z}_i = (z_{1,i_1}^\top, z_{2,i_2}^\top)^\top$, with $z_{j,i_j} = (y_{j,i_j}^\top, x_{j,i_j}^\top)^\top$; further set $\mathbf{z}_i = (\mathbf{y}_i^\top, \mathbf{x}_i^\top)^\top$, with $\mathbf{y}_i = (y_{1,i_1}^\top, y_{2,i_2}^\top)^\top$ and $\mathbf{x}_i = (x_{1,i_1}^\top, x_{2,i_2}^\top)^\top$, and let $\mathbf{Z} = (\mathbf{Y}^\top, \mathbf{X}^\top)^\top$ be an i.i.d. copy of it. Denote $D_n(t) = \{\mathbf{i} : 1 \leq i_j \leq n_{jt}, j = 1, 2\}$ for the collection of indices for the observations at stage t , and $D_n = \cup_t D_n(t)$. Although in practice the number of stages is at least 2 and no more than 10, no such restriction is needed for theoretical developments; in fact, one can even consider t as varying in a continuous interval.

To formulate the proposed U-statistic with covariate information, let

$$w_i = F(\{\mathbf{z}_i\}) = F_1(\{z_{1,i_1}\})F_2(\{z_{2,i_2}\})$$

be the weight assigned to the observed data z_{1,i_1} and z_{2,i_2} and $w(t) = \{w_i(t) : \mathbf{i} \in D_n(t)\}$. Given that the $w_i(t)$'s are unknown (as is F), we maximize the product of the $w_i(t)$'s subject to appropriate constraints (they may not be independent of each other) to get the empirical weights as below. We define $\bar{g}(\mathbf{x}_i) = \{g(x_{1,i_1}) + g(x_{2,i_2})\}/2$. Clearly, $E\{\bar{g}(\mathbf{x}_i)\} = \mathbf{0}$. Rewrite the EL subject to the auxiliary information constraints at stage t as maximizing

$$\max_{w(t)} \prod_{\mathbf{i} \in D_n(t)} w_i(t) \quad (1)$$

subject to

$$\sum_{\mathbf{i} \in D_n(t)} w_i(t) = 1 \quad \text{and} \quad \sum_{\mathbf{i} \in D_n(t)} w_i(t) \bar{g}(\mathbf{x}_i) = \mathbf{0}.$$

We get, as in Owen [18],

$$w_i(t) = \frac{1}{n_{1,t} n_{2,t}} \frac{1}{1 + \lambda^\top(t) \bar{g}(\mathbf{x}_i)},$$

and $\lambda(t) = \lambda_n(t) = (\lambda_{n,1}(t), \dots, \lambda_{n,d}(t))^\top$ where, for all $j \in \{1, \dots, r\}$, $\lambda_{n,j}(t) = \lambda_{n,j}(t)(\mathbf{x}_1, \dots, \mathbf{x}_n)$ is determined by the following equation:

$$\sum_{\mathbf{i} \in D_n(t)} \frac{\bar{g}(\mathbf{x}_i)}{1 + \lambda^\top(t) \bar{g}(\mathbf{x}_i)} = \mathbf{0}.$$

For details regarding the existence of $\lambda(t)$ as the solution of the above equation, see, e.g., the papers by Owen and others. The above proposed weights for U-statistics are parallel to those in the EL, and simpler than existing methods in that there is no need to form a product of m elements from w_1, \dots, w_n as in Wood et al. [27], nor to merge the data as in Jing et al. [8].

With the weights given above, we construct the weighted two-sample U-statistic estimate of θ , at stage t , as

$$U_n(t) = \sum_{\mathbf{i} \in D_n(t)} w_i(t) h(\mathbf{y}_i). \quad (2)$$

Note that in the absence of covariates, the rank statistics in Huang et al. [5], Huang et al. [6], Huang and Tan [4] and other related ones can all be expressed as U-statistics. Thus our statistic (2) can be viewed as a generalization of these statistics with covariate information incorporated. It is known that the (one-sample) U-statistic is the minimal variance unbiased estimator of the kernel mean; see, e.g., Serfling [22, p. 176]. Now with the covariate information incorporated, using the conditioning (on the covariate) argument, it can be shown that the two-sample U-statistic given in (2) is also the minimal variance unbiased estimator of the kernel mean.

3.2. Asymptotic properties

Before investigating the large-sample properties of the proposed U-statistics, we review some basic terminologies, notations and facts needed for the study of large-sample properties of multi-sample U-statistics. The number of times a variable appears in the kernel is called the multiplicity of this variable. Suppose an m -sample kernel $h(y_{ji}, i = 1, \dots, k_j, j = 1, \dots, m)$ is symmetric within each sample. In our specific case, $m = 2, k_1 = k_2 = 1$, and

$$h(y_1, y_2) = \sum_{j=1}^k a_j \{ \mathbf{1}(y_{1j} < y_{2j}) - \mathbf{1}(y_{1j} > y_{2j}) \}.$$

For general examples, we refer the readers to the book of Koroljuk and Borovskich [11], hereafter KB, (Sections 1.2 and 4.5). The parameter of interest is

$$\theta = E_F \{ h(Y_{ji}, i = 1, \dots, k_j, j = 1, \dots, m) \} < \infty,$$

where $F = (F_1, \dots, F_m)$, and F_j is the distribution of the j th sample. For $n = (n_1, \dots, n_m)$, the m -sample U-statistic with kernel h and multiplicity $\mathbf{k} = (k_1, \dots, k_m)$ is

$$U_n = \prod_{j=1}^m \frac{1}{\binom{n_j}{k_j}} \sum_{\{1 \leq i_1 < \dots < i_{k_j} \leq n_j, j=1, \dots, m\}} h(y_{jr} : r = i_{j_1}, \dots, i_{j_{k_j}}, j = 1, \dots, m).$$

As in the pioneering work of Hoeffding [2,3] for one-sample U-statistics, the multi-sample U-statistic U_n can be written as a linear combination of canonical forms $\tilde{h}_{\mathbf{r}}$; see, e.g., KB (p. 38).

For a vector $\mathbf{r} = (r_1, \dots, r_k)$ of integers, denote $|\mathbf{r}| = r_1 + \dots + r_k$. If $\tilde{h}_{\mathbf{r}} = 0$ for all \mathbf{r} with $|\mathbf{r}| < s$ and $\tilde{h}_{\mathbf{r}} \neq 0$ for some \mathbf{r} with $|\mathbf{r}| = s$, the U-statistic U_n with kernel h is said to be of rank s , where $1 \leq s \leq |\mathbf{k}|$. When $s > 1$, U_n is called *degenerate*; in this case the convergence rate is not \sqrt{n} and the asymptotic distribution is not Gaussian.

For vectors $\mathbf{k} = (k_1, \dots, k_m)$ and $\mathbf{r} = (r_1, \dots, r_m)$ of integers, the notation $\mathbf{r} \leq \mathbf{k}$ means component-wise less or equal. Let

$$\eta_{\mathbf{r}}^2 = E\{\tilde{h}_{\mathbf{r}}^2(Y_{ji} : 1 \leq i \leq r_j, j = 1, \dots, m)\}.$$

If $\text{rank}(h) = s$, then $\eta_{\mathbf{r}} = 0$ for all \mathbf{r} with $|\mathbf{r}| < s$. The following variance formula holds (KB, p. 38), with $\binom{k}{0} = 1$ and $\eta_0^2 = 0$,

$$\text{var}(U_n) = \prod_{j=1}^m \frac{1}{\binom{n_j}{k_j}} \sum_{\{\mathbf{r} \leq \mathbf{k}\}} \prod_{j=1}^m \binom{k_j}{r_j} \binom{n_j - k_j}{k_j - r_j} \eta_{\mathbf{r}}^2.$$

For our two-sample U-statistic defined in (2) with the given symmetric kernel h , the multiplicity is $(k_1, k_2) = (1, 1)$, $\bar{\theta} = E\{h(Y_1, Y_2)\}$, $\text{rank}(h) = 1$ and h is non-degenerate. In this case,

$$\begin{aligned} \tilde{h}_{(1,0)}(y_{1,i_1}) &= E\{h(y_{1,i_1}, Y_2)\} - \bar{\theta}, \\ \tilde{h}_{(0,1)}(y_{2,i_2}) &= E\{h(Y_1, y_{2,i_2})\} - \bar{\theta}, \\ \tilde{h}_{(1,1)}(y_{1,i_1}, y_{2,i_2}) &= h(y_{1,i_1}, y_{2,i_2}) - \bar{\theta} - \tilde{h}_{(1,0)}(y_{1,i_1}) - \tilde{h}_{(0,1)}(y_{2,i_2}). \end{aligned}$$

When there is no covariate, or $g \equiv \mathbf{0}$, or $w_i = n_{1,t}^{-1}n_{2,t}^{-1}$, we have

$$\text{var}\{U_n(t)\} = \frac{1}{n_{1,t}n_{2,t}} \{ (n_{2,t} - 1)\eta_{(1,0)}^2 + (n_{1,t} - 1)\eta_{(0,1)}^2 + \eta_{(1,1)}^2 \}.$$

With covariate information incorporated, we will see later that the asymptotic variance of $U_n(t)$ can often be reduced.

Similarly, the function \tilde{g} is a d -dimensional symmetric kernel with multiplicity $(1,1)$, and it has canonical forms $\tilde{g}_{\mathbf{r}} = (\tilde{g}_{\mathbf{r},1}, \dots, \tilde{g}_{\mathbf{r},d})^\top$ for $\mathbf{r} = (1,0), (0,1)$ and $(1,1)$. The canonical forms $\tilde{h}_{\mathbf{r}}$ and $\tilde{g}_{\mathbf{r}}$ exist theoretically, but are unknown in practice since F is unknown. Let

$$r_0 = \min\{\text{rank}(\tilde{g}_1), \dots, \text{rank}(\tilde{g}_d)\}, \quad r_1 = \min\{\text{rank}(\tilde{g}_1 h), \dots, \text{rank}(\tilde{g}_d h)\}.$$

With the mapping h , we will see that for most practical problems, we have $r_1 = r_0 = 1$, i.e., the corresponding kernels are non-degenerate.

To study the problem generally, below we will consider the sample sizes n_{jt} at interim times t in a more general setting. We consider t can vary in an interval $[t_0, T]$ with $0 < t_0 < T < \infty$, and let $N = \sup_{t \in [t_0, T]} \max(n_{1,t}, n_{2,t})$. Denote $F = (F_1, F_2)$, and note that $\bar{\theta} = \bar{\theta}(F)$ is a functional of F . Let $\mathcal{F} = \{F\}$ be the collection of all distribution functions of $Z = (Y, X)$. Further denote $\|a\|^\alpha = (a_1^2 + \dots + a_d^2)^{\alpha/2}$. The following conditions will be used in this section:

(C1) $\Omega = E\{g(\mathbf{X})g^\top(\mathbf{X})\}$ is positive definite.

(C2) $E\|g(\mathbf{X})\|^\alpha < \infty$ for some $\alpha > 0$ to be specified.

(C3) X has bounded support and g is continuous almost everywhere for all $F \in \mathcal{F}$.

Let $n_t = \sqrt{n_{1,t}n_{2,t}}$. As the weights w_i generally have no closed form solution, we first show how they can be approximated. In what follows, $\mathbf{1}_d = (1, \dots, 1)^\top$ is a d -dimensional vector of 1's.

Lemma. Assume (C1) and (C2) for $\alpha > 2d/r_0$, and $\min(n_{1,t}, n_{2,t}) \rightarrow \infty$. One then has the following results, in which the $O(\cdot)$ terms are uniformly over the \mathbf{x}_i .

(i)

$$w_i(t) \stackrel{a.s.}{=} \frac{1}{n_{1,t}n_{2,t}} \left[1 - \bar{g}^\top(\mathbf{x}_i) 2\Omega^{-1} \frac{1}{n_{1,t}n_{2,t}} \sum_{j \in D_n(t)} \bar{g}(\mathbf{x}_j) \right. \\ \left. + \mathbf{1}_d^\top \bar{g}(\mathbf{x}_i) O\{\rho_n n_t^{-1/2} (\ln \ln n_t)^{1/2}\} + \{\mathbf{1}_d^\top \bar{g}(\mathbf{x}_i) + \|g_+(\mathbf{x}_i)\|^2\} O(\rho_n^2) \right],$$

where

$$\rho_n = \begin{cases} O\{n_t^{-1/2} (\ln \ln n_t)^{1/2}\}, & r_0 = 1; \\ o(n_t^{-r_0/2} \ln n_t), & 1 < r_0 \leq d. \end{cases}$$

(ii)

$$w_i(t) = \frac{1}{n_{1,t}n_{2,t}} \left[1 - \bar{g}^\top(\mathbf{x}_i) 2\Omega^{-1} \frac{1}{n_{1,t}n_{2,t}} \sum_{j \in D_n(t)} \bar{g}(\mathbf{x}_j) \right. \\ \left. + \mathbf{1}_d^\top \bar{g}(\mathbf{x}_i) O_p\{n_t^{-(r_0+1)/2}\} + \{\mathbf{1}_d^\top \bar{g}(\mathbf{x}_i) + \|\bar{g}(\mathbf{x}_i)\|^2\} O_p(n_t^{-r_0}) \right].$$

The following results characterize the consistency, exact convergence rate, and uniform consistency of the estimator $U_n(t)$ of $\bar{\theta}$.

Theorem 1. Under the conditions of the Lemma, one has

- (i) $U_n(t) \rightarrow \bar{\theta}$ almost surely.
- (ii) With $\sigma^2(t)$ given in Theorem 2(i) below, one has almost surely

$$\limsup_{n_t} \left\{ 2\sigma^2(t) \frac{\ln \ln n_t}{n_t} \right\}^{-1/2} |U_n(t) - \bar{\theta}| = 1.$$

- (iii) If (C3) also holds, then $\sup_{F \in \mathcal{F}} \sup_{t \in [t_0, T]} |U_n(t) - \bar{\theta}(F)| \rightarrow 0$ almost surely.

Let \rightsquigarrow stands for weak convergence in $\ell^\infty[t_0, T]$, the space of all bounded real functions on $[t_0, T]$ equipped with the sup metric. The following results describe the asymptotic distribution of $\{U_n(t)\}$ as a sequence of point estimators and of processes on $[t_0, T]$.

Theorem 2. Assume conditions of the Lemma, and that $n_{2,t}/n_{1,t} \rightarrow c(t)$ with $0 < c(t) < \infty$. Then

- (i) $\sqrt{n_t} \{U_n(t) - \bar{\theta}\} \rightsquigarrow \mathcal{N}[0, \sigma^2(t)]$, where

$$\sigma^2(t) = \sum_{j=1}^2 c^{3/2-j}(t) \tau_j^2$$

with $\mathbf{r}_1 = (1, 0)$, $\mathbf{r}_2 = (0, 1)$, and for $j \in \{1, 2\}$,

$$\tau_j^2 = \begin{cases} \eta_{\mathbf{r}_j}^2 - 2A^\top \Omega^{-1} A_{\mathbf{r}_j} + A^\top \Omega^{-1} A, & \text{if } r_0 = 1, \\ \eta_{\mathbf{r}_j}^2, & \text{if } r_0 > 1, \end{cases}$$

with $\eta_{\mathbf{r}_j}^2 = E_F\{\tilde{h}_{\mathbf{r}_j}^2(Y_j)\}$, $A_{\mathbf{r}_j} = E_F\{g(X_1)\tilde{h}_{\mathbf{r}_j}(Y_1)\}$, and $A = E_F\{\bar{g}(\mathbf{X})h(\mathbf{Y})\} = (A_{(1,0)} + A_{(0,1)})/2$.

- (ii) Denote $\mathbb{U}_n(t) = \sqrt{n_t} \{U_n(t) - \bar{\theta}(F)\}$ and, for $s \leq t$, denote $b(s, t) = \lim n_{1,s}/n_{1,t}$. If (C3) also holds, then $\mathbb{U}_n \rightsquigarrow \mathbb{Z}$ in $\ell^\infty[t_0, T]$, where \mathbb{Z} is a mean zero Gaussian process on $[t_0, T]$ with covariance function defined, for all $s, t \in [t_0, T]$, by

$$R(s, t) = E\{\mathbb{Z}(s)\mathbb{Z}(t)\} = c^{1/4}(s \wedge t) b^{1/2}(s \wedge t, s \vee t) \sum_{j=1}^2 c^{5/4-j}(s \vee t) \tau_j^2.$$

(iii) Assume conditions of (ii), under $\mathcal{H}'_0 : \bar{\theta} = \mathbf{0}$, then

$$\sup_{t \in [t_0, T]} \sqrt{tn_t/(T - t_0)} \sigma^{-1}(t) U_n(t) \rightsquigarrow W, \quad \inf_{t \in [t_0, T]} \sqrt{tn_t/(T - t_0)} \sigma^{-1}(t) U_n(t) \rightsquigarrow -W$$

where W has distribution function defined, for all $w \geq 0$, by

$$G(w) = \frac{2}{\sqrt{2\pi}} \int_0^w e^{-x^2/2} dx.$$

From Theorem 2 we see that the most interesting case is $r_0 = r_1 = 1$, in which $\sqrt{n_t}\{U_n(t) - \theta\}$ is asymptotic non-degenerate normal, with asymptotic variance being smaller than that without covariates, or $g \equiv \mathbf{0}$. With covariates incorporated, the deduction in the asymptotic variance is $\sum_{j=1}^2 c^{3/2-j}(t) A_{\mathbf{r}_j}^\top \Omega^{-1} A_{\mathbf{r}_j}$, so the estimate of $\bar{\theta}$ is more accurate than the case without covariates. Also with the covariate information implemented, the estimate $U_n(t)$ of $\bar{\theta}$ is efficient in the sense that its asymptotic variance achieves the information lower bound; see Theorem 3 in Yuan et al. [29] for the one-sample case. We will see later that for most practical problems we have $r_0 = r_1 = 1$.

For the given kernel h , we have

$$\begin{aligned} \tilde{h}_{(1,0)}(\mathbf{Y}_1) &= \sum_{j=1}^k a_j \{1 - 2F_{2j}(Y_{1j})\} - E\{h(\mathbf{Y}_1, \mathbf{Y}_2)\}, \\ \tilde{h}_{(0,1)}(\mathbf{Y}_2) &= \sum_{j=1}^k a_j \{2F_{1j}(Y_{2j}) - 1\} - E\{h(\mathbf{Y}_1, \mathbf{Y}_2)\}, \end{aligned}$$

where F_{2j} is the marginal distribution function of Y_{2j} , and F_{1j} is that of Y_{1j} . For $g(x) = x - \mu$, we have $r_0 = 1$, $\Omega = \text{cov}(X_1) = \text{cov}(X_2)$, $\tilde{g}(X) = (X_1 + X_2)/2 - \mu$, $\tilde{g}_{(1,0)}(X_1) = (X_1 - \mu)/2$, $\tilde{g}_{(0,1)}(X_2) = (X_2 - \mu)/2$.

3.3. Empirical confidence interval for $\bar{\theta}$

A confidence interval for $\bar{\theta}$ can be constructed using the asymptotic Gaussian distribution given in Theorem 2. However, a more natural confidence interval is given by the empirical likelihood ratio, in which the interval is not necessarily symmetric around $\bar{\theta}$. Below we give such a result.

Let $G(\mathbf{z}|\bar{\theta}) = (\bar{g}^\top(\mathbf{x}), h(\mathbf{y}) - \bar{\theta})^\top$. We have $E_F\{G(\mathbf{Z}|\bar{\theta})\} = \mathbf{0}$. Without auxiliary information, the weights that maximize (1) subject to $\sum_{\mathbf{i} \in D_n(t)} w_{\mathbf{i}} = 1$ are $w_{\mathbf{i}} = (n_{1,t}n_{2,t})^{-1}$ for all $\mathbf{i} \in D_n(t)$; while the weights that maximize (1) subject to $\sum_{\mathbf{i} \in D_n(t)} w_{\mathbf{i}} = 1$ and $\sum_{\mathbf{i} \in D_n(t)} w_{\mathbf{i}} G(\mathbf{Z}_{\mathbf{i}}|\bar{\theta}) = \mathbf{0}$ are $w_{\mathbf{i}} = (n_{1,t}n_{2,t})^{-1} / \{1 + \lambda^\top G(\mathbf{Z}_{\mathbf{i}}|\bar{\theta})\}$ and λ is determined by (2) with g_+ replaced by $G(\cdot|\bar{\theta})$. Therefore we define the empirical log likelihood ratio of $\bar{\theta}$ with the presence of auxiliary information by

$$R_G(\bar{\theta}) = \frac{L_n(\bar{\theta})}{(n_{1,t}n_{2,t})^{n_{1,t}n_{2,t}}} = \prod_{\mathbf{i} \in D_n(t)} (n_{1,t}n_{2,t}w_{\mathbf{i}}),$$

where

$$L_n(\bar{\theta}) = \max_{\substack{\sum_{\mathbf{i} \in D_n(t)} w_{\mathbf{i}} = 1, \\ \sum_{\mathbf{i} \in D_n(t)} w_{\mathbf{i}} G(\mathbf{Z}_{\mathbf{i}}|\bar{\theta}) = \mathbf{0}}} \prod_{\mathbf{i} \in D_n(t)} w_{\mathbf{i}}$$

and denote

$$l(\bar{\theta}) = -\ln R_G(\bar{\theta}) = \sum_{\mathbf{i} \in D_n(t)} \ln\{1 + \lambda^\top G(\mathbf{Z}_{\mathbf{i}}|\bar{\theta})\}.$$

Let

$$\Lambda = E_F\{G(\mathbf{Z}|\bar{\theta})G'(\mathbf{Z}|\bar{\theta})\} = \begin{pmatrix} \Omega & A \\ A^\top & \eta^2 \end{pmatrix},$$

$\eta^2 = \text{var}\{h(\mathbf{Z})\}$, $\Lambda_1 = \text{cov}(\tilde{G}_1)$, and \tilde{G}_1 be the first canonical form (vector) of G .

Note that when there is no auxiliary information, $G(\cdot|\bar{\theta})$ reduces to $h(\cdot) - \bar{\theta}$, and λ is a scalar determined by

$$\sum_{\mathbf{i} \in D_n(t)} \{h(\mathbf{y}_{\mathbf{i}}) - \bar{\theta}\} / [1 + \lambda\{h(\mathbf{y}_{\mathbf{i}}) - \bar{\theta}\}] = 0.$$

The corresponding log-likelihood ratio is

$$\ell_h(\bar{\theta}) = \sum_{\mathbf{i} \in D_n(t)} \ln[1 + \lambda\{h(\mathbf{y}_{\mathbf{i}}) - \bar{\theta}\}].$$

Theorem 3. Under conditions of Theorem 2(i), assume $r_0 = 1$ and Λ to be positive definite. Then

$$\frac{n_t}{n_{1,t}n_{2,t}} \ell(\bar{\theta}) \rightsquigarrow Z_{d+1}^\top \Lambda_1^{1/2} \Lambda^{-1} \Lambda_1^{1/2} Z_{d+1}, \quad Z_{d+1} \sim \mathcal{N}(\mathbf{0}, I_{d+1}).$$

Furthermore, $n_t \eta^2 \ell_h(\bar{\theta}) / \{n_{1,t}n_{2,t}\eta_1^2\} \rightsquigarrow \chi_1^2$.

For a given level α , let $G^{-1}(1 - \alpha)$ be the $(1 - \alpha)$ th upper quantile of the asymptotic distribution in Theorem 3, the approximate level α confidence region of $\bar{\theta}$ is given by

$$\left\{ \bar{\theta} : \frac{n_t}{n_{1,t}n_{2,t}} \ell(\bar{\theta}) \leq G^{-1}(1 - \alpha) \right\},$$

where $G^{-1}(1 - \alpha)$ can be computed by simulation.

When $\Lambda_1^{1/2} = \Lambda^{1/2}$, the above result for U-statistic automatically reduces to that for the common EL ratio, and the first limit in Theorem 3 is χ_{d+1}^2 ; see the corresponding result in Theorem 2 of Qin and Lawless [20]. Therefore, with covariate auxiliary information incorporated in the likelihood ratio, the length of the confidence region for $\bar{\theta}$ cannot be reduced. This is an interesting contrast to the estimation with auxiliary information, in which the asymptotic variance is reduced. However, using the EL ratio, the shape of the confidence region is more natural than many other commonly used methods, such as the normal approximation, which are forced to be symmetric. The latter method may have poorer coverage probability because of the shorter interval length and its subjective shape.

4. Stopping boundaries

A group sequential clinical trial is determined by its stopping boundaries, which serve as guidelines for clinicians to monitor the developments in the trial process. In each stage, if the test statistic falls within the boundary, the trial is continued to the next stage, otherwise it is terminated. There are different ways to construct the stopping boundary; common choices include the O'Brien–Fleming [17] boundary and the sequential conditional probability ratio tests (SCPRT) boundary [23,28]. We will use the SCPRT boundary, as this method provides monitoring procedures with the appealing property that a decision reached at early stopping would be maintained with high probability should the trial continue to its planned end. For this we need to formulate the estimator $U_n(t)$ as a Brownian motion as in Xiong, Tan and Boyett [28] and Huang and Tan [4], and check the basic conditions as in the following corollary. Let $D(n_{1,t}, n_{2,t}) = \sqrt{n_{1,t}n_{2,t}} U_n(t)/\sigma(t)$, and $I(n_{1,t}, n_{2,t}) = n_t$ be the information time.

Corollary 1. Under the conditions of Theorem 2, the process $\{D(n_{1,t}, n_{2,t}) : 0 < t \leq 1\}$ converges weakly to a Brownian motion with drift $\delta(t) = \bar{\theta}/\sigma(t)$ at information times $\{I(n_{1,t}, n_{2,t}) : 0 < t \leq 1\}$, i.e., for fixed t , $I^{-1/2}(n_{1,t}, n_{2,t})D(n_{1,t}, n_{2,t})$ is asymptotically Gaussian, and

$$\begin{aligned} E\{D(n_{1,t}, n_{2,t})\} &= \delta(t)I(n_{1,t}, n_{2,t}) + o(1), & \text{var}\{D(n_{1,t}, n_{2,t})\} &= I(n_{1,t}, n_{2,t}) + o(1), \\ \text{cov}\{D(n_{1,s}, n_{2,s}), D(n_{1,t}, n_{2,t})\} &= \text{var}\{D(n_{1,s}, n_{2,s})\} + o(1) \quad \text{for any } 0 < s \leq t \leq 1. \end{aligned}$$

Now we return to the testing of $\mathcal{H}_0 : \bar{\theta} \leq 0$ vs. $\mathcal{H}_1 : \bar{\theta} > 0$, which has the same rejection rule as testing $\mathcal{H}'_0 : \bar{\theta} = 0$ vs. \mathcal{H}_1 . Define the test statistic as $T_n(t) = \sqrt{n_t} U_n(t)$. By Theorem 2(i), the level α test of \mathcal{H}'_0 vs. \mathcal{H}_1 is given by the rejection rule: $T_n(t) > \sigma(t)\Phi^{-1}(1 - \alpha)$, where Φ^{-1} is the quantile function of the standard Normal distribution.

Theorem 2(iii) can be used as a reference for the stopping/continuation of the trial. If

$$\sup_{t \in [t_0, T]} \sqrt{tn_t/(T - t_0)} \sigma^{-1}(t) U_n(t) < G^{-1}(1 - \alpha)$$

or

$$\max_{t \in \{t_1, \dots, t_m\}} \sqrt{tn_t/(T - t_0)} \sigma^{-1}(t) U_n(t) < G^{-1}(1 - \alpha)$$

and if

$$\inf_{t \in [t_0, T]} \sqrt{tn_t/(T - t_0)} \sigma^{-1}(t) U_n(t) > G^{-1}(1 - \alpha)$$

or

$$\min_{t \in \{t_1, \dots, t_m\}} \sqrt{tn_t/(T - t_0)} \sigma^{-1}(t) U_n(t) > G^{-1}(1 - \alpha),$$

then it suggests that, at level α , either the two treatments have no significant difference, or the trial should go beyond time T to detect noticeable difference.

4.1. Stopping boundaries

The stopping boundaries for each stage should satisfy the condition of family-wise error rate no greater than α , for some specified α . There are some commonly used stopping boundaries, including the O'Brien–Fleming (OBF) boundary (O'Brien and Fleming [17]) and the sequential conditional probability ratio test (SCPRT) boundary considered, e.g., in Xiong, Tan and Boyett [28] and (Huang and Tan [4]). The SCPRT stopping boundaries for a k -stage clinical trial are given, for all $r \in \{1, \dots, k\}$, by

$$a_r = -z_{\alpha/2} - \{2as_r(1 - s_r)\}^{1/2}, \quad b_r = z_{\alpha/2} + \{2bs_r(1 - s_r)\}^{1/2},$$

where $z_{\alpha/2}$ is the upper quantile of order $1 - \alpha/2$ of the standard Normal distribution. Often a symmetric boundary pair is used, and value(s) of $a = b$, depend on the discordance probability ρ , given in Table 1 in Xiong, Tan and Boyett [28]; $s_r = \text{var}\{V_n(t_r)\}/\text{var}\{V_n(T)\}$ is the time information fraction, and $V_n(t) = n_{1,t}n_{2,t}U_n(t)$. By Theorem 2 we have asymptotically

$$s_r = \frac{\sigma^2(t_r)}{\sigma^2(T)} \left(\frac{n_{1,t_r}n_{2,t_r}}{n_{1,T}n_{2,T}} \right)^{3/2},$$

with $\sigma^2(t)$ given in Theorem 2, and $n_{1,T}$ and $n_{2,T}$ are the sample sizes of the two treatment groups at the final stage. As an example, if the trial has $k = 4$ stages and we choose $\rho = 0.02$, then from Table 1 in Xiong, Tan and Boyett [28] we find $a (=b) = 2.953$; if we choose $\rho = 0.005$, then $a (=b) = 4.227$; and the lower/upper boundaries a_r/b_r are obtained by the above formula.

4.2. Sequential stopping rule

In applications, in each interim trial stage r , we compute the statistic $T_n(t_r) = \sqrt{n_{t_r}} U_n(t_r)$, if its value is within the boundaries, i.e., if $T_n(t_r) = \sqrt{n_{t_r}} U_n(t_r) \in (a_r, b_r)$, we continue to the next stage ($r + 1$), otherwise we stop the clinical trial and make the corresponding decision. In this case the computations of type I error and power are different from those of the single stage trial; below we address these issues.

4.3. Type I error

Recall that in the single stage case, the theoretic (or asymptotic) type I error is the significance level α , while the observed type I error in the simulation is, for the one-sided hypothesis,

$$\alpha = \Pr_{\bar{\theta}=0} \left\{ \sqrt{n_{t_1}} \sigma^{-1}(t_1) U_n(t_1) > b_1 \right\} = \Pr_{\bar{\theta}=0} \left\{ \sqrt{n_{t_1}} \sigma^{-1}(t_1) U_n(t_1) > z_\alpha \right\}.$$

In the above we used the fact that when $k = 1$, $t_1 = T$, $s_1 = 1$ and so $b_1 = z_\alpha$.

In the general $k (> 1)$ stages case, we need to control the family-wise type I error to be no more than α , and for continuous type statistic, to be equal α i.e.,

$$\alpha = \sum_{r=1}^k \Pr_{\bar{\theta}=0} \left\{ \sqrt{n_{t_j}} \sigma^{-1}(t_j) U_n(t_j) \in (a_j, b_j), j < r; \sqrt{n_{t_r}} \sigma^{-1}(t_r) U_n(t_r) > b_r \right\}.$$

The type I error in the group sequential case for a two-sided hypothesis is obtained similarly, with $a_r = 0$ and b_r as before but with z_α replaced by $z_{\alpha/2}$.

4.4. Power

Recall in the single stage case, the theoretic (or asymptotic) power for given $\bar{\theta} \neq 0$ can be computed from the asymptotic distribution, while the observed power in the simulation is, for the one-sided hypothesis,

$$\beta = \beta(\bar{\theta}) = \Pr_{\bar{\theta}} \left\{ \sqrt{n_{t_1}} \sigma^{-1}(t_1) U_n(t_1) > b_1 \right\} = \Pr_{\bar{\theta}} \left\{ \sqrt{n_{t_1}} \sigma^{-1}(t_1) U_n(t_1) > z_\alpha \right\}.$$

In the general $k (> 1)$ stages case,

$$\beta = \beta(\bar{\theta}) = \sum_{r=1}^k \Pr_{\bar{\theta}} \left\{ \sqrt{n_{t_j}} \sigma^{-1}(t_j) U_n(t_j) \in (a_j, b_j), j < r; \sqrt{n_{t_r}} \sigma^{-1}(t_r) U_n(t_r) > b_r \right\}.$$

The power in the group sequential case for a one-sided hypothesis is obtained similarly, with $a_r = 0$ and b_r as before but with z_α replaced by $z_{\alpha/2}$.

Table 1Estimated $\bar{\theta}$ by two estimators (with estimated variances).

(n_1, n_2)	(30, 30)	(50, 50)	(80, 100)	(100, 150)
$U_{n,0}$	0.1725926 (0.3536608)	0.1824000 (0.3266168)	0.2322500 (0.3080994)	0.2625778 (0.3139383)
U_n	0.1162940 (0.2491390)	0.1630525 (0.2731421)	0.2267121 (0.2074212)	0.2503937 (0.2834820)

Table 2

Overall type I error and power for U-statistics with and without covariates (normal error distribution).

$\bar{\theta}$	k_y	k_x	R	$n_{r1}(=n_{r2})$	With covariates				Without covariates			
					Type I error $\alpha = 0.05$		Power $1 - \beta \geq 0.8$		Type I error $\alpha = 0.05$		Power $1 - \beta \geq 0.8$	
					SCPRT	OBF	SCPRT	OBF	SCPRT	OBF	SCPRT	OBF
0.223	2	3	1	(100)	0.050	0.050	0.880	0.880	0.048	0.046	0.870	0.876
			2	(50, 100)	0.054	0.054	0.826	0.806	0.050	0.050	0.820	0.806
			3	(33, 67, 100)	0.052	0.052	0.874	0.836	0.048	0.048	0.872	0.830
			4	(25, 50, 75, 100)	0.044	0.054	0.866	0.854	0.044	0.052	0.866	0.862
0.116	2	5	1	(300)	0.052	0.046	0.816	0.818	0.050	0.046	0.816	0.816
			2	(150, 300)	0.046	0.050	0.806	0.806	0.044	0.050	0.804	0.800
			3	(100, 200, 300)	0.050	0.046	0.850	0.822	0.050	0.042	0.842	0.818
			4	(75, 150, 225, 300)	0.048	0.054	0.816	0.808	0.040	0.052	0.818	0.802
0.304	3	3	1	(150)	0.050	0.048	0.844	0.842	0.044	0.046	0.842	0.846
			2	(75, 150)	0.052	0.052	0.840	0.852	0.050	0.050	0.836	0.850
			3	(50, 100, 150)	0.052	0.052	0.880	0.806	0.052	0.048	0.884	0.814
			4	(38, 75, 112, 150)	0.048	0.048	0.856	0.820	0.048	0.046	0.858	0.826
0.156	3	5	1	(300)	0.046	0.046	0.812	0.846	0.046	0.046	0.808	0.846
			2	(150, 300)	0.052	0.042	0.892	0.868	0.052	0.042	0.882	0.864
			3	(100, 200, 300)	0.044	0.048	0.864	0.866	0.044	0.046	0.856	0.866
			4	(75, 150, 225, 300)	0.054	0.054	0.848	0.854	0.054	0.050	0.844	0.850

Notes: $\bar{\theta}$: global treatment effect (GTE), k_y : number of endpoints, k_x : number of Covariates; R : total number of stages; $n_{r1}(=n_{r2})$: cumulative sample size which is assumed to be equal in both groups at stage r .

5. Simulation studies and real data analysis

5.1. Simulation studies

Let $k = \dim(y) = 3$, $\mathbf{a} = (1/3, 1/3, 1/3)^\top$, so $\bar{\theta}$ is the Mann–Whitney difference parameter. We first generate the covariates X_{1i} 's and X_{2i} 's from a 5-dimensional Gaussian distribution with the same mean vector $\mu_x = (3.1, 1.8, -1.5, 2.2, 0.3)^\top$ and a given covariance matrix Ω . The two samples Y_{1i} and Y_{2j} 's are generated from a multivariate regression model given the covariates, i.e., for every $i \in \{1, \dots, n_1\}$ and $j \in \{1, \dots, n_2\}$,

$$Y_{1i} = \beta X_{1i} + \mu_1 + \epsilon_{1i}, \quad Y_{2j} = \beta X_{2j} + \mu_2 + \epsilon_{2j},$$

where $\beta = (-0.38, 0.64, 1.13, -0.32, 1.13)^\top$, $\mu_1 = (1.5, 2.0, 0.7)^\top$ and $\mu_2 = (0.4, 0.7, -1.4)^\top$, and the error terms ϵ_{1i} and ϵ_{2j} are i.i.d. $\mathcal{N}(0, 1)$. The latter choice of distribution is for convenience only; the method applies for any error distribution, as the U-statistic is a nonparametric method.

The true value of θ is $\bar{\theta}_0 = 0.2278281$. We estimate $\bar{\theta}_0$ by two versions of two-sample U-statistics, $U_{n,0}$ without covariate, and U_n with the covariates incorporated, with $\bar{g}(\mathbf{x}_i) = (X_{1,i} + X_{2i})/2 - \mu_x$. The results in Table 1 confirm that the variance of the U-statistics with covariate incorporated is smaller, thus is more efficient, as expected. As one can see, the amount of variance deduction depends on the nature of the covariates.

The type I error can be simulated as follows. Suppose there are $k = 3$ stages, with sample size n_1 , n_2 and n_3 at each stage. For simplicity, assume that the test is one-sided, for fixed level α , with boundaries C_1 , C_2 , and C_3 ; the total number of repetitions is $m = 1000$. At the j th repetition, we need to sample $n = n_1 + n_2 + n_3$ data under \mathcal{H}_0 and compute the test statistic $\sqrt{n_j} U_n(t_j)$. Let r_j be the rejection status for the j th repetition, i.e., $r_j = 1$ if the j th repetition is a rejection, otherwise $r_j = 0$. To compute r_j , first sample n_1 data and compute $T_{j1} = \sqrt{n_1} U_n(t_1)$. If $T_{j1} \geq C_1$, set $r_j = 1$ and the j th repetition is terminated; otherwise go to the second stage with additional n_2 samples and compute $T_{j2} = \sqrt{n_1 + n_2} U_n(t_2)$. If $T_{j2} \geq C_2$, set $r_j = 1$ and the j th repetition is terminated; otherwise go to the third stage with an additional n_3 samples and compute $T_{j3} = \sqrt{n} U_n(t_3)$. If $T_{j3} \geq C_3$, set $r_j = 1$; otherwise set $r_j = 0$. The simulated type I error is then given by $(r_1 + \dots + r_m)/m$. To estimate the power, one can proceed the same way except that the data are generated under \mathcal{H}_1 .

Results of a simulation are reported in Table 2. It shows that both the overall Type I error and power are reasonable, i.e., similar to those of the test from a single stage trial.

Table 3

Parkinson disease trial group sequential U-statistics based test results (2 covariates: age and years of education).

R	r	(n _{r1} , n _{r2})	$\bar{\theta}$	t	SCPRT				OBF		
					Test stat	a _t	b _t	dscn	Test stat	c _t	dscn
Analysis based on 6-month improvement with total sample size 646											
1	1	(348, 298)	0.2249	1.00	7.5107	1.64	1.64	Reject	7.5107	1.64	Reject
2	1	(174, 149)	0.1845	0.50	3.0813	−0.20	1.85	Reject	4.3576	2.42	Reject
	2	(348, 298)	0.2249	1.00	7.5107	1.64	1.64	Reject	7.5107	1.71	Reject
3	1	(116, 99)	0.1836	0.33	2.0404	−0.54	1.63	Reject	3.5370	3.00	Reject
	2	(232, 199)	0.2186	0.67	4.8710	0.01	2.18	Reject	5.9632	2.12	Reject
	3	(348, 298)	0.2249	1.00	7.5107	1.64	1.64	Reject	7.5107	1.73	Reject
4	1	(87, 74)	0.1730	0.25	1.4385	−0.64	1.46	Continue	2.8819	3.50	Continue
	2	(174, 149)	0.1845	0.50	3.0813	−0.39	2.04	Reject	4.3576	2.48	Reject
	3	(261, 224)	0.2137	0.75	5.3587	0.18	2.29	Reject	6.1843	2.02	Reject
	4	(348, 298)	0.2249	1.00	7.5107	1.64	1.64	Reject	7.5107	1.75	Reject
Analysis based on 9-month improvement with total sample size 572											
1	1	(322, 250)	0.1817	1.00	5.5581	1.64	1.64	Reject	5.5581	1.64	Reject
2	1	(161, 125)	0.1206	0.50	1.8448	−0.20	1.85	Continue	2.6090	2.42	Reject
	2	(322, 250)	0.1817	1.00	5.5581	1.64	1.64	Reject	5.5581	1.71	Reject
3	1	(107, 83)	0.1230	0.33	1.2494	−0.54	1.63	Continue	2.1678	3.00	Continue
	2	(215, 167)	0.1729	0.67	3.5311	0.01	2.18	Reject	4.3208	2.12	Reject
	3	(322, 250)	0.1817	1.00	5.5581	1.64	1.64	Reject	5.5581	1.73	Reject
4	1	(80, 62)	0.1379	0.25	1.0468	−0.64	1.46	Continue	2.1010	3.50	Continue
	2	(161, 125)	0.1206	0.50	1.8448	−0.39	2.04	Continue	2.6090	2.48	Reject
	3	(242, 188)	0.1621	0.75	3.7269	0.18	2.29	Reject	4.2983	2.02	Reject
	4	(322, 250)	0.1817	1.00	5.5581	1.64	1.64	Reject	5.5581	1.75	Reject

Notes: R: total number of stages, r: stage, $\bar{\theta}$: global treatment effect, t: information fraction; (n_{r1}, n_{r2}) : cumulative sample size for treatment and control group at stage r;

test stat: test statistics based on U-statistics;

 (a_t, b_t) : lower and upper SCPRT boundary; c_t : OBF boundary, dscn: Decision at each stage.

5.2. Real data analysis

The National Institutes of Health (NIH) sponsored the Deprenyl and Tocopherol Antioxidative Therapy of Parkinsonism (DATATOP), a multi-center for randomized controlled clinical trial for studying the neuroprotective treatment. The Parkinson Study Group [24] conducted this DATATOP from September 1987 to November 1989 to investigate whether the initiation of levodopa therapy, necessitated by disability for patients with early untreated Parkinson's Disease (PD), could be postponed by long-term use of antioxidative agents such as deprenyl or tocopherol. There were 800 patients with early Parkinson's Disease enrolled in this trial at 28 sites across the United States and Canada, and they were randomly assigned to one of four treatment groups: (1) active deprenyl, (2) active tocopherol, (3) active deprenyl and tocopherol, and (4) placebo. The development of disability requiring the onset of levodopa therapy was the primary endpoint in the DATATOP trial. The Parkinson Study Group reported the results of both the tocopherol and deprenyl treatments after a mean follow-up time of 14 ± 6 months. No tocopherol effect was found in the study, while deprenyl had a strong beneficial effect which significantly delayed the need of levodopa therapy, and slowed the disease progression measured by the total Unified Parkinson's Disease Rating Scale (UPDRS), its subscales about motor, mentation and activities of daily living. Moreover, some covariates such as baseline age, years of education, race and gender could be potential confounders that are associated with the deprenyl treatment and the disease progression.

In this paper, three key movement dysfunction measures from UPDRS subscales (motor, mentation and activities of daily living) and four covariates information, such as baseline age, years of education, race and gender, were retrospectively analyzed to see whether treatment benefit of deprenyl on these endpoints could be identified earlier with a smaller sample size, rather than until the planned end of the study when data from all 800 patients had been collected. If so, it would amount to considerable human and financial savings.

Recall the one-sided hypothesis. The proposed group sequential U-statistics based on the test was applied to the data on responses at 6 months and 9 months in motor, mentation and activities of daily living, with baseline covariates age, years of education, race and gender. All observations from endpoints were coded such that larger values correspond to better clinical conditions. The endpoints measured by UPDRS and its sub-scales are ordinal with higher numbers indicating more severe disease. To test the beneficial effect of deprenyl treatment, it is necessary to flip the order of observations from endpoints by adding a negative sign when coding for analysis.

With the proposed group sequential test, Tables 3 and 4 suggest a pronounced early treatment effect since the first or second stage data already demonstrated the significance for both the 6-month and the 9-month analysis: the stochastic process hits the early stopping upper bound for efficacy, and the chance of a potential reversion of the conclusion is negligible.

Table 4

Parkinson disease trial group sequential U-statistics based test results (4 covariates: age, years of education, race, gender).

R	r	(n _{r1} , n _{r2})	$\bar{\theta}$	t	SCPRT				OBF		
					Test stat	a _t	b _t	dscn	Test stat	c _t	dscn
Analysis based on 6-month improvement with total sample size 646											
1	1	(348,298)	0.2250	1.00	7.5274	1.64	1.64	Reject	7.5274	1.64	Reject
2	1	(174,149)	0.1829	0.50	3.0603	−0.20	1.85	Reject	4.3280	2.42	Reject
	2	(348,298)	0.2250	1.00	7.5274	1.64	1.64	Reject	7.5274	1.71	Reject
3	1	(116,99)	0.1831	0.33	2.0385	−0.54	1.63	Reject	3.5337	3.00	Reject
	2	(232,199)	0.2183	0.67	4.8747	0.01	2.18	Reject	5.9678	2.12	Reject
	3	(348,298)	0.2250	1.00	7.5274	1.64	1.64	Reject	7.5274	1.73	Reject
4	1	(87,74)	0.1723	0.25	1.4359	−0.64	1.46	Continue	2.8766	3.50	Continue
	2	(174,149)	0.1829	0.50	3.0603	−0.39	2.04	Reject	4.3279	2.48	Reject
	3	(261,224)	0.2140	0.75	5.3771	0.18	2.29	Reject	6.2054	2.02	Reject
	4	(348,298)	0.2250	1.00	7.5274	1.64	1.64	Reject	7.5274	1.75	Reject
Analysis based on 9-month improvement with total sample size 572											
1	1	(322,250)	0.1815	1.00	5.5792	1.64	1.64	Reject	5.5792	1.64	Reject
2	1	(161,125)	0.1199	0.50	1.8425	−0.20	1.85	Continue	2.6056	2.42	Reject
	2	(322,250)	0.1815	1.00	5.5792	1.64	1.64	Reject	5.5792	1.71	Reject
3	1	(107,83)	0.1339	0.33	1.3667	−0.54	1.63	Continue	2.3715	3.00	Continue
	2	(215,167)	0.1687	0.67	3.4619	0.01	2.18	Reject	4.2362	2.12	Reject
	3	(322,250)	0.1815	1.00	5.5792	1.64	1.64	Reject	5.5792	1.73	Reject
4	1	(80,62)	0.1614	0.25	1.2310	−0.64	1.46	Continue	2.4709	3.50	Continue
	2	(161,125)	0.1199	0.50	1.8425	−0.39	2.04	Continue	2.6056	2.48	Reject
	3	(242,188)	0.1586	0.75	3.6647	0.18	2.29	Reject	4.2266	2.02	Reject
	4	(322,250)	0.1815	1.00	5.5792	1.64	1.64	Reject	5.5792	1.75	Reject

Notes: R: total number of stages, r: stage, $\bar{\theta}$: global treatment effect, t: information fraction; (n_{r1}, n_{r2}) : cumulative sample size for treatment and control group at stage r;

test stat: test statistics based on U-statistics;

 (a_t, b_t) : lower and upper SCPRT boundary; c_t : OBF boundary, dscn: Decision at each stage.

Table 5 shows the result for the same data without covariate information. We see that the decisions with and without covariates are very similar, the only difference is for the 6-month improvement data, at stage 1 of the 4-stage trial with the SCPRT boundary. In this case the method with covariates information incorporated (Tables 3 and 4) suggests continuing to the next stage trial, while the method without covariate information (Table 5) suggests discontinue.

6. Concluding remarks

In this paper, we presented a group sequential test statistic that adjusts for potential confounding baseline covariates for treatment comparison with multiple primary mixed endpoints for multi-stage clinical trial design and analysis. The test was derived using a two-sample U-statistic incorporating auxiliary information from the covariates. We provided the large-sample distribution properties of our new test and a method to compute the sequential stopping boundaries.

The proposed test is suitable to use to solve the Behrens–Fisher problem when the assumption of equal covariance structure between the two comparing groups is violated. Simulation studies have shown that this new group sequential test for multi-dimensional endpoints can well preserve the desired operating characteristics (such as overall Type I error and power). When we applied this new test to analyze a real Parkinson disease trial data, we were able to conclude the significant treatment beneficial effect at an early stage as compared to the trial investigators who reached the same conclusion only at the end of the trial. This cannot only shorten the trial duration but also greatly reduce the trial cost.

Our work can be further extended to include the identification of endpoints with stronger effects and the potential dropping of irrelevant ones from the evaluation. Although the U-statistic with auxiliary information from covariates has an advantage that they have smaller asymptotic variance compared to that without auxiliary information, the resulting estimates can be biased if incorporating incorrect auxiliary information about covariates (Yuan et al. [29]). Therefore, it is paramount to justify the covariate auxiliary information properly.

Acknowledgments

We appreciate the help and suggestions from the Editor and Associate Editors to improve the presentation of the paper.

Appendix

Proof of the Lemma. The proof is tedious but standard and hence omitted. It will be provided upon request. \square

Table 5
Parkinson disease trial group sequential U-statistics based test results (no covariates).

R	r	(n_{r1}, n_{r2})	$\bar{\theta}$	t	SCPRT				OBF		
					Test stat	a_t	b_t	dscn	Test stat	c_t	dscn
Analysis based on 6-month improvement with total sample size 646											
1	1	(348,298)	0.2250	1.00	7.4968	1.64	1.64	Reject	7.4968	1.64	Reject
2	1	(174,149)	0.1853	0.50	3.0881	−0.20	1.85	Reject	4.3672	2.42	Reject
	2	(348,298)	0.2250	1.00	7.4968	1.64	1.64	Reject	7.4968	1.71	Reject
3	1	(116,99)	0.1915	0.33	2.1227	−0.54	1.63	Reject	3.6797	3.00	Reject
	2	(232,199)	0.2185	0.67	4.8593	0.01	2.18	Reject	5.9490	2.12	Reject
	3	(348,298)	0.2250	1.00	7.4968	1.64	1.64	Reject	7.4968	1.73	Reject
4	1	(87,74)	0.1923	0.25	1.5960	−0.64	1.46	Reject	3.1974	3.50	Continue
	2	(174,149)	0.1853	0.50	3.0881	−0.39	2.04	Reject	4.3672	2.48	Reject
	3	(261,224)	0.2138	0.75	5.3496	0.18	2.29	Reject	6.1737	2.02	Reject
	4	(348,298)	0.2250	1.00	7.4968	1.64	1.64	Reject	7.4968	1.75	Reject
Analysis based on 9-month improvement with total sample size 572											
1	1	(322,250)	0.1818	1.00	5.5511	1.64	1.64	Reject	5.5511	1.64	Reject
2	1	(161,125)	0.1168	0.50	1.7826	−0.20	1.85	Continue	2.5210	2.42	Reject
	2	(322,250)	0.1818	1.00	5.5511	1.64	1.64	Reject	5.5511	1.71	Reject
3	1	(107,83)	0.1223	0.33	1.2406	−0.54	1.63	Continue	2.1526	3.00	Continue
	2	(215,167)	0.1709	0.67	3.4857	0.01	2.18	Reject	4.2653	2.12	Reject
	3	(322,250)	0.1818	1.00	5.5511	1.64	1.64	Reject	5.5511	1.73	Reject
4	1	(80,62)	0.1458	0.25	1.1053	−0.64	1.46	Continue	2.2185	3.50	Continue
	2	(161,125)	0.1168	0.50	1.7826	−0.39	2.04	Continue	2.5210	2.48	Reject
	3	(242,188)	0.1611	0.75	3.6989	0.18	2.29	Reject	4.2661	2.02	Reject
	4	(322,250)	0.1818	1.00	5.5511	1.64	1.64	Reject	5.5511	1.75	Reject

Notes: R: total number of stages, r: stage, $\bar{\theta}$: global treatment effect, t: information fraction;
 (n_{r1}, n_{r2}) : cumulative sample size for treatment and control group at stage r;
test stat: test statistics based on U-statistics;
 (a_t, b_t) : lower and upper SCPRT boundary; c_t : OBF boundary, dscn: Decision at each stage.

Proof of Theorem 1. Part (i): Let $U_{0,n}(t) = \sum_{i \in D_n(t)} h(\mathbf{y}_i) / (n_{1,t} n_{2,t})$. By Lemma (i), we have, almost surely,

$$U_n(t) = U_{0,n}(t) - \left\{ \frac{1}{n_{1,t} n_{2,t}} \sum_{i \in D_n(t)} \bar{g}^\top(\mathbf{x}_i) h(\mathbf{y}_i) \right\} 2\Omega^{-1} \left\{ \frac{1}{n_{1,t} n_{2,t}} \sum_{j \in D_n(t)} \bar{g}(\mathbf{x}_j) \right\} \\ + O\{\rho_n n_t^{-1/2} (\ln \ln n_t)^{1/2}\} \frac{1}{n_{1,t} n_{2,t}} \sum_{i \in D_n(t)} 1_d^\top \bar{g}(\mathbf{x}_i) h(\mathbf{y}_i) + O(\rho_n^2) \frac{1}{n_{1,t} n_{2,t}} \sum_{i \in D_n(t)} \{1_d^\top \bar{g}(\mathbf{x}_i) + \|\bar{g}(\mathbf{x}_i)\|^2\} h(\mathbf{y}_i).$$

By the given conditions and the SLLN of two-sample U-statistics, $U_{0,n}(t) \xrightarrow{\text{a.s.}} \bar{\theta}$,

$$U_{1,n}(t) = \frac{1}{n_{1,t} n_{2,t}} \sum_{i \in D_n(t)} \bar{g}(\mathbf{x}_i) \xrightarrow{\text{a.s.}} E_F\{\bar{g}(\mathbf{X})\} = 0,$$

$$U_{2,n}(t) = \frac{1}{n_{1,t} n_{2,t}} \sum_{i \in D_n(t)} \bar{g}(\mathbf{x}_i) h(\mathbf{y}_i) \xrightarrow{\text{a.s.}} A < \infty,$$

where $A = E_F\{\bar{g}(\mathbf{X}) h(\mathbf{Y})\}$, and

$$U_{3,n}(t) = \frac{1}{n_{1,t} n_{2,t}} \sum_{i \in D_n(t)} \{1_d^\top \bar{g}(\mathbf{x}_i) + \|\bar{g}(\mathbf{x}_i)\|^2\} h(\mathbf{y}_i) \\ \xrightarrow{\text{a.s.}} E_F[\{1_d^\top \bar{g}(\mathbf{X}) + \|\bar{g}(\mathbf{X})\|^2\} h(\mathbf{Y})] < \infty.$$

Thus we have, almost surely,

$$U_n(t) - \bar{\theta} = U_{0,n}(t) - \bar{\theta} - O(1)U_{1,n}(t) + O\{\rho_n n_t^{-1/2} (\ln \ln n_t)^{1/2}\} + O(\rho_n^2) \rightarrow 0.$$

Part (ii): Using Lemma (i) and notations in the proof of (i), we have, almost surely,

$$U_n(t) - \bar{\theta} = \frac{1}{n_{1,t} n_{2,t}} \sum_{i \in D_n(t)} \{h(\mathbf{y}_i) - \bar{\theta} - A^\top 2\Omega^{-1} \bar{g}(\mathbf{x}_i)\} - \{U_{2,n}(t) - A\}^\top 2\Omega^{-1} U_{1,n}(t) \\ + O\{\rho_n n_t^{-1/2} (\ln \ln n_t)^{1/2}\} 1_d^\top U_{2,n}(t) + O(\rho_n^2) U_{3,n}(t).$$

By the Strong Law of Large Numbers for two-sample U-statistics, $U_{3,n} \rightarrow C_3$ almost surely for some $C_3 < \infty$.

Note that the corresponding results in Chapter 3 of KB for one-sample U-statistics actually holds for two-sample U-statistics, and so Corollary 3.4.1 there holds for our case with $\binom{n}{r}$ replaced by $n_{1,t}n_{2,t}$. Thus we have $U_{1,n}(t) = o(n^{-q})$ (a.s.) and $U_{2,n}(t) - A = o(n^{-q})$ (a.s.) for all $q < 1/2$, and so, a.s.

$$U_n(t) - \bar{\theta} = \frac{1}{n_{1,t}n_{2,t}} \sum_{\mathbf{i} \in D_n(t)} \{h(\mathbf{y}_i) - \bar{\theta} - A^\top 2\Omega^{-1} \bar{g}(\mathbf{x}_i)\} + o(n^{-2q}) + O\{\rho_n n^{-1/2} (\ln \ln n)^{1/2}\} + O(\rho_n^2).$$

Therefore, Theorem 9.1.1 of KB also holds for two-sample U-statistics, i.e., the Law of the Iterated Logarithm holds for the first term above, these give the desired result.

Part (iii): Using the expression of $U_n(t) - \bar{\theta}(F)$ at the beginning of (ii), we first prove that

$$\sup_{F \in \mathcal{F}} \sup_{t \in [t_0, T]} \left| \frac{1}{n_{1,t}n_{2,t}} \sum_{\mathbf{i} \in D_n(t)} \{h(\mathbf{y}_i) - \bar{\theta}(F) - A^\top 2\Omega^{-1} \bar{g}(\mathbf{x}_i)\} \right| \rightarrow 0 \quad (\text{A.1})$$

holds almost surely. Then the almost sure convergence to 0 of $\sup_{F \in \mathcal{F}} \sup_{t \in [t_0, T]} |U_{1,n}(t)|$, $\sup_{F \in \mathcal{F}} \sup_{t \in [t_0, T]} |U_{2,n}(t) - A|$, and $\sup_{F \in \mathcal{F}} \sup_{t \in [t_0, T]} |U_{3,n}(t)|$ will follow the same way, and the desired result will be proved.

Now we prove (A.1). Denote $n_j = \sup_{t \in [t_0, T]} n_{j,t}$, and rewrite the average in (A.1) as

$$\frac{n_1 n_2}{n_{1,0} n_{2,0}} \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \{h(y_{1,i}, y_{2,j}) - \bar{\theta}(F) - A^\top 2\Omega^{-1} \bar{g}(x_{1,i}, x_{2,j})\} \mathbf{1}(i \leq n_{1,t}) \mathbf{1}(j \leq n_{2,t}).$$

In the above summation, the summands are not i.i.d. due to the indicator functions. However, the expression is asymptotically equivalent to the following i.i.d. summation

$$\frac{n_1 n_2}{n_{1,0} n_{2,0}} \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \{h(y_{1,i}, y_{2,j}) - \bar{\theta}(F) - A^\top 2\Omega^{-1} \bar{g}(x_{1,i}, x_{2,j})\} v_{1,i} v_{2,j},$$

where the $v_{1,i} = v_{1,i}(t)$'s are i.i.d, the $v_{2,j} = v_{2,j}(t)$'s are i.i.d, with

$$\Pr(v_{1,i} = 1) = \frac{n_{1,t}}{n_1} = 1 - \Pr(v_{1,i} = 0), \quad \Pr(v_{2,i} = 1) = \frac{n_{2,t}}{n_2} = 1 - \Pr(v_{2,j} = 0).$$

Since $\lim(n_1 n_2)/(n_{1,0} n_{2,0}) < \infty$, to prove (A.1) we only need to prove

$$\sup_{F \in \mathcal{F}} \sup_{t \in [t_0, T]} \left| \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \{h(\mathbf{y}_i) - \bar{\theta}(F) - A^\top 2\Omega^{-1} \bar{g}(\mathbf{x}_i)\} v_{1,i} v_{2,j} \right| \rightarrow 0 \quad (\text{a.s.}) \quad (\text{A.2})$$

For each fixed (y, x) , let $\zeta(y, x|F) = E_F\{h(Y_1, y) - A^\top 2^{-1} \Omega^{-1} \bar{g}(X_1, x)\}$. Then $E_F\{\zeta(Y_2, X_2|F)\} = \bar{\theta}(F)$. Let

$$\left| \frac{1}{n_1} \sum_{i=1}^{n_1} \{h(y_{1,i}, y) - \zeta(y, x|F) - A^\top 2\Omega^{-1} \bar{g}(x_{1,i}, x)\} v_{1,i} \right| = \left| \frac{1}{n_1} \sum_{i=1}^{n_1} a(y_{1i}, x_{1i}, y, x|F) \right| = d_{n_1}(y, x|F),$$

and for fixed F and t , let $\mathcal{A}_0 = \{a(\cdot, \cdot, y, x|F) : (y, x) \in \mathbb{R}^{k+d}\}$. We are to prove that \mathcal{A}_0 is a Glivenko–Cantelli class uniformly for $F \in \mathcal{F}$. However, the functions $a(\cdot, \cdot, y, x|F)$ in \mathcal{A}_0 depend on the $n_{j,t}$'s through the $v_{1,i}$ and $v_{2,j}$, which is not easy to deal with. So instead we work on the set of functions $\mathcal{A} = \{a(\cdot, \cdot, y, x|F) : (y, x) \in \mathbb{R}^{k+d}, \text{ with } v_{1,i}, v_{2,j} \text{ replaced by all } u_{1,i}(t), u_{2,j}(t)\text{'s, with } 0 \leq u_{1,i}(t), u_{2,j}(t) \leq 1, t \in [t_0, T]\}$. Obviously, $\mathcal{A}_0 \subset \mathcal{A}$, and we prove that \mathcal{A} is a uniform Glivenko–Cantelli class on \mathcal{F} , so that \mathcal{A}_0 is as well.

Note that h and g_+ is fixed, and by (C3) g is almost everywhere bounded continuous. Hence for any $\epsilon > 0$, the bracketing covering number for \mathcal{A} is $N_{[\cdot]}(\epsilon, \mathcal{A}, L_1(F)) < \infty$, so by Theorem 2.4.1 in van der Vaart and Wellner [26, p. 122], hereafter VW, for each fixed F and t , \mathcal{A} is a Glivenko–Cantelli class for F at t , i.e., one has almost surely

$$\sup_{(y,x) \in \mathbb{R}^{k+d}} d_{n_1}(y, x|F) \rightarrow 0.$$

Since $a(\cdot, \cdot, y, x|F)$ is a fixed function, $|h(\cdot)| \leq 1$ and $|\zeta(\cdot|F)| \leq 1$, $q(\cdot) = 2 + |A^\top \{g(\cdot) + g(x)\}|$ is an envelope function for \mathcal{A} . By (C3) we have

$$\lim_{M \rightarrow \infty} \sup_{F \in \mathcal{F}} \sup_{t \in [t_0, T]} E_F[q(\mathbf{X}) \mathbf{1}\{q(\mathbf{X}) > M\}] = 0.$$

The above is the first condition of Theorem 2.8.1 in VW. Also, let $N(\epsilon, \mathcal{A}, \|\cdot\|)$ be the ϵ covering number of \mathcal{A} without bracketing, under the norm $\|\cdot\|$, and $L_1(Q)$ be the L_1 norm under probability measure Q , $\|q\|_{Q,1} = \int |q| dQ$. With (C3), and the relationship between $N(\epsilon, \mathcal{A}, \|\cdot\|)$ and $N_{[\cdot]}(\epsilon, \mathcal{A}, \|\cdot\|)$ (VW, p. 84), we have

$$\sup_Q \ln N(\epsilon \|q\|_{Q,1}, \mathcal{A}, L_1(Q)) \leq \sup_Q \ln N_{[\cdot]}(2\epsilon \|q\|_{Q,1}, \mathcal{A}, L_1(Q)) < \infty,$$

where the sup is taken over all probability measures Q on X . The above implies the second condition in Theorem 2.8.1 in VW, and so by this theorem, \mathcal{A} is a uniform Glivenko–Cantelli class for F and t , i.e., one has, almost surely,

$$\sup_{t \in [t_0, T]} \sup_{(y, x) \in \mathbb{R}^{k+d}} \sup_{F \in \mathcal{F}} d_{n_1}(y, x|F) \rightarrow 0.$$

The expression in (A.2) is bounded above by

$$\frac{1}{n_2} \sum_{i=1}^{n_2} \left| \frac{1}{n_1} \sum_{j=1}^{n_1} \{h(y_{1,i}, y_{2,j}) - \zeta(y_{2,i}, x_{2,j}|F) - A^\top 2\Omega^{-1} \bar{g}(x_{1,i}, x_{2,j})\} v_{1,i} \right| + \left| \frac{1}{n_2} \sum_{j=1}^{n_2} \{\zeta(y_{2,j}, x_{2,j}|F) - \bar{\theta}(F)\} v_{2,j} \right|,$$

which can be rewritten as

$$\frac{1}{n_2} \sum_{j=1}^{n_2} d_{n_1}(y_{2,j}, x_{2,j}|F) + \left| \frac{1}{n_2} \sum_{j=1}^{n_2} \{\zeta(y_{2,j}, x_{2,j}|F) - \bar{\theta}(F)\} v_{2,j} \right|.$$

By the result on $d_{n_1}(y, x|F)$ just obtained, the first term above is uniformly small over the $x_{2,j}$'s and \mathcal{F} for large n_2 . Since $\zeta(\cdot|F)$ is formed by the fixed bounded function h , and $\bar{\theta}(F) = E_F\{\zeta(Y_2, X_2|F)\}$ in the same way as in the proof for $d_{n_1}(y, x|F)$, we have, almost surely,

$$\sup_{t \in [t_0, T]} \sup_{y \in \mathbb{R}^k} \sup_{F \in \mathcal{F}} \left| \frac{1}{n_2} \sum_{j=1}^{n_2} \{\zeta(y_{2,j}, x_{2,j}|F) - \bar{\theta}(F)\} v_{2,j} \right| \rightarrow 0.$$

This gives (A.2) and completes the proof. \square

Proof of Theorem 2. Part (i): Using the fact that $U_{1,n}(t) \rightarrow A$ and $U_{2,n}(t) \rightarrow C_2$ almost surely for some $C_2 < \infty$ as proved in Theorem 1(i), it follows from Lemma (ii) that

$$\begin{aligned} \sqrt{n_t} \{U_n(t) - \bar{\theta}\} &= \sqrt{n_t} \frac{1}{n_{1,t} n_{2,t}} \sum_{\mathbf{i} \in D_n(t)} \{h(\mathbf{y}_i) - \bar{\theta} - A^\top 2\Omega^{-1} \bar{g}(\mathbf{x}_i)\} \\ &\quad - \sqrt{n_t} \{U_{2,n}(t) - A\}^\top 2\Omega^{-1} U_{1,n}(t) + O_p(n_t^{-r_0/2}) 1_d^\top U_{2,n}(t) + O_p(n_t^{-(r_0-1/2)}) U_{3,n}(t). \end{aligned}$$

The second term on the right-hand side above is, for all $q \in (0, 1/2)$, $n_t^{1/2} O(n_t^{-2q}) = o_p(1)$; the third term above is $O_p(n_t^{-r_0/2})$ as $U_{2,n}(t) \rightarrow A < \infty$ almost surely; and the last term above is $O_p(n_t^{-(r_0-1/2)})$ as $U_{3,n}(t) \rightarrow C_3$ almost surely for some $C_3 < \infty$; Thus we only need to deal with the first term on the right-hand side above.

Recall that $\tilde{g}_{(1,0)}$ and $\tilde{g}_{(0,1)}$ are the canonical forms of g . Let $\tilde{\tilde{g}}_{(1,0)}$ and $\tilde{\tilde{g}}_{(0,1)}$ be the canonical forms of \bar{g} , and

$$H(\mathbf{y}) = h(\mathbf{y}) - \tilde{h}_{(1,0)}(y_1) - \tilde{h}_{(0,1)}(y_2) - \bar{\theta}, \quad G(\mathbf{x}) = \bar{g}(\mathbf{x}) - \tilde{\tilde{g}}_{(1,0)}(x_1) - \tilde{\tilde{g}}_{(0,1)}(x_2).$$

Then $\mathcal{H}_1(y_1) = E\{H(\mathbf{Y})|Y_1 = y_1\} = 0$, i.e., $H(\mathbf{y})$ is a degenerate kernel. Similarly, $G(\mathbf{x})$ is degenerate, so is $K(\mathbf{z}) = H(\mathbf{y}) - A^\top 2\Omega^{-1} G(\mathbf{x})$, with $E_F\{K(\mathbf{Z})\} = 0$ and so $r_k = \text{rank}(K) \geq 2$. Now we have

$$\begin{aligned} \sqrt{n_t} \{U_n(t) - \bar{\theta}\} &= \sqrt{n_t} \frac{1}{n_{1,t} n_{2,t}} \sum_{\mathbf{i} \in D_n(t)} \left[\tilde{h}_{(1,0)}(y_{1,i_1}) + \tilde{h}_{(0,1)}(y_{2,i_2}) - A^\top 2\Omega^{-1} \{\tilde{\tilde{g}}_{(1,0)}(x_{1,i_1}) + \tilde{\tilde{g}}_{(0,1)}(x_{2,i_2})\} \right] \\ &\quad + \sqrt{n_t} \frac{1}{n_{1,t} n_{2,t}} \sum_{\mathbf{i} \in D_n(t)} K(\mathbf{z}_i) + O_p(n_t^{-1/2}) + o_p(1), \end{aligned}$$

which can be rewritten as

$$\begin{aligned} \sqrt{n_t} &\left[\frac{1}{n_{1,t}} \sum_{i=1}^{n_{1,t}} \left\{ \tilde{h}_{(1,0)}(y_{1i}) - A^\top 2\Omega^{-1} \tilde{\tilde{g}}_{(1,0)}(x_{1i}) \right\} + \frac{1}{n_{2,t}} \sum_{i=1}^{n_{2,t}} \left\{ \tilde{h}_{(0,1)}(y_{2i}) - A^\top 2\Omega^{-1} \tilde{\tilde{g}}_{(0,1)}(x_{2i}) \right\} \right] \\ &+ \sqrt{n_t} \frac{1}{n_{1,t} n_{2,t}} \sum_{\mathbf{i} \in D_n(t)} K(\mathbf{z}_i) + O_p(n_t^{-1/2}) + o_p(1). \end{aligned}$$

Let \tilde{K}_r be the canonical forms of K , and $\xi_r^2 = E_F\{\tilde{K}_r^2(\mathbf{Z})\} < \infty$ by the given conditions, and with $|\mathbf{r}| \geq 2$. Given that K is degenerate, $\xi_r^2 = 0$ for $|\mathbf{r}| = 1$ and $\xi_r^2 \geq 0$ for $|\mathbf{r}| = 2$. Note $|\mathbf{r}|$ cannot be greater than 2 as the multiplicity of K is $\mathbf{k} = (1, 1)$ with $|\mathbf{k}| = 2$. So by the variance formula for two-sample U-statistics with multiplicity $(1, 1)$ as given in KB (p. 38),

$$\begin{aligned} \text{var} \left\{ \sqrt{n_t} \frac{1}{n_{1,t} n_{2,t}} \sum_{\mathbf{i} \in D_n(t)} K(\mathbf{z}_i) \right\} &= \frac{n_t}{n_{1,t} n_{2,t}} \binom{1}{1} \binom{1}{1} \binom{n_{1,t}-1}{0} \binom{n_{2,t}-1}{0} \xi_{(1,1)}^2 \\ &= O\{(n_{1,t} n_{2,t})^{-1/2}\} \rightarrow 0, \end{aligned}$$

and so $\sqrt{n_t} (n_{1,t} n_{2,t})^{-1} \sum_{\mathbf{i} \in D_n(t)} K(\mathbf{z}_i) \xrightarrow{P} 0$. Recalling that $n_{2,t} = c(t)n_{1,t}$, we get, with $A = E\{\tilde{g}(\mathbf{X})h(\mathbf{Y})\}$,

$$\begin{aligned} \sqrt{n_t} \{U_n(t) - \bar{\theta}\} &= \frac{c^{1/4}(t)}{\sqrt{n_{1,t}}} \sum_{i=1}^{n_{1,t}} \{\tilde{h}_{(1,0)}(y_{1i}) - A^\top 2\Omega^{-1} \tilde{g}_{(1,0)}(x_{1i})\} \\ &\quad + \frac{c^{-1/4}(t)}{\sqrt{n_{2,t}}} \sum_{i=1}^{n_{2,t}} \{\tilde{h}_{(0,1)}(y_{2i}) - A^\top 2\Omega^{-1} \tilde{g}_{(0,1)}(x_{2i})\} + o_P(1). \end{aligned}$$

Note that $\tilde{g}_{(1,0)}(x_{1i}) = g(x_{1i})/2$ and $\tilde{g}_{(0,1)}(x_{2i}) = g(x_{2i})/2$, and the (y_{1i}, x_{1i}) 's and (y_{2i}, x_{2i}) 's are independent. Therefore, we get

$$\sqrt{n_t} \{U_n(t) - \bar{\theta}\} = \frac{c^{1/4}(t)}{\sqrt{n_{1,t}}} \sum_{i=1}^{n_{1,t}} \{\tilde{h}_{(1,0)}(y_{1i}) - A^\top \Omega^{-1} g(x_{1i})\} + \frac{c^{-1/4}(t)}{\sqrt{n_{2,t}}} \sum_{i=1}^{n_{2,t}} \{\tilde{h}_{(0,1)}(y_{2i}) - A^\top \Omega^{-1} g(x_{2i})\} + o_P(1).$$

Let $A_{(1,0)} = E\{g(X_1)\tilde{h}_{(1,0)}(Y_1)\}$ and $A_{(0,1)} = E\{g(X_2)\tilde{h}_{(0,1)}(Y_2)\}$. Note that $\text{rank}(h) = 1$, and

$$\begin{aligned} A &= E\{\tilde{g}(\mathbf{X})h(\mathbf{Y})\} = \frac{1}{2} \left(E[g(X_1)\{\tilde{h}_{(1,0)}(Y_1) + \bar{\theta}\}] + E[g(X_2)\{\tilde{h}_{(0,1)}(Y_2) + \bar{\theta}\}] \right) \\ &= \frac{1}{2} \left(E\{g(X_1)\tilde{h}_{(1,0)}(Y_1)\} + E\{g(X_2)\tilde{h}_{(0,1)}(Y_2)\} \right) = (A_{(1,0)} + A_{(0,1)})/2. \end{aligned}$$

If $r_0 = \min\{\text{rank}(g_1), \dots, \text{rank}(g_d)\} = 1$, then

$$\begin{aligned} \text{var}\{\tilde{h}_{(1,0)}(Y_1) - A^\top \Omega^{-1} g(X_1)\} &= \eta_{(1,0)}^2 - 2A_{(1,0)}^\top \Omega^{-1} A_{(1,0)} + A_{(1,0)}^\top \Omega^{-1} A_{(1,0)} \\ &= \eta_{(1,0)}^2 - \frac{3}{4} A_{(1,0)}^\top \Omega^{-1} A_{(1,0)} - \frac{1}{2} A_{(1,0)}^\top \Omega^{-1} A_{(0,1)} + \frac{1}{4} A_{(0,1)}^\top \Omega^{-1} A_{(0,1)}. \end{aligned}$$

Similarly,

$$\begin{aligned} \text{var}\{\tilde{h}_{(0,1)}(Y_2) - A^\top \Omega^{-1} g(X_2)\} &= \eta_{(0,1)}^2 - 2A_{(0,1)}^\top \Omega^{-1} A_{(0,1)} + A_{(0,1)}^\top \Omega^{-1} A_{(0,1)} \\ &= \eta_{(0,1)}^2 - \frac{3}{4} A_{(0,1)}^\top \Omega^{-1} A_{(0,1)} - \frac{1}{2} A_{(0,1)}^\top \Omega^{-1} A_{(1,0)} + \frac{1}{4} A_{(1,0)}^\top \Omega^{-1} A_{(1,0)}. \end{aligned}$$

If $r_0 > 1$, then $\tilde{g}_{(1,0)} = \tilde{g}_{(0,1)} \equiv 0$, and

$$\text{var}\{\tilde{h}_{(1,0)}(Y_1) - A^\top 2\Omega^{-1} \tilde{g}_{(1,0)}(X_1)\} = \text{var}\{\tilde{h}_{(1,0)}(Y_1)\} = \eta_{(1,0)}^2$$

while $\text{var}\{\tilde{h}_{(0,1)}(Y_2) - A^\top 2\Omega^{-1} \tilde{g}_{(0,1)}(X_2)\} = \eta_{(0,1)}^2$. With these and the Central Limit Theorem, now we get $\sqrt{n_t} \{U_n(t) - \bar{\theta}\} \rightsquigarrow \mathcal{N}[0, \sigma^2(t)]$ with $\sigma^2(t)$ given in Theorem 2(i). Part (ii): Using the expression at the beginning of (i), and the same way as in the proof of Theorem 1(iii), we have

$$\sup_{F \in \mathcal{F}} \sup_{t \in [t_0, T]} |U_{1,n}(t)| \xrightarrow{\text{a.s.}} 0, \quad \sup_{F \in \mathcal{F}} \sup_{t \in [t_0, T]} |U_{2,n}(t) - A| \xrightarrow{\text{a.s.}} 0, \quad \sup_{F \in \mathcal{F}} \sup_{t \in [t_0, T]} |U_{3,n}(t) - C_3| \xrightarrow{\text{a.s.}} 0,$$

and so we only need to prove

$$\sup_{F \in \mathcal{F}} \sup_{t \in [t_0, T]} \left| \sqrt{n_t} \frac{1}{n_{1,t} n_{2,t}} \sum_{\mathbf{i} \in D_n(t)} \{h(\mathbf{y}_i) - \bar{\theta} - A^\top 2^{-1} \Omega^{-1} g_+(\mathbf{x}_i)\} - \mathbb{Z}(t) \right| \xrightarrow{P} 0. \quad (\text{A.3})$$

Let $\mathbf{r}_1 = (1, 0)$ and $\mathbf{r}_2 = (0, 1)$. The left-hand side above is bounded above by

$$\begin{aligned} &\sup_{F \in \mathcal{F}} \sup_{t \in [t_0, T]} \left| \sqrt{n_t} \frac{1}{n_{1,t}} \sum_{i=1}^{n_{1,t}} \{\tilde{h}_{(1,0)}(y_{1i}) - A^\top 2^{-1} \Omega^{-1} \tilde{g}_{+, (1,0)}(x_{1i})\} - \mathbb{Z}_1(t) \right| \\ &\quad + \sup_{F \in \mathcal{F}} \sup_{t \in [t_0, T]} \left| \sqrt{n_t} \frac{1}{n_{2,t}} \sum_{i=1}^{n_{2,t}} [\tilde{h}_{(0,1)}(y_{2i}) - A^\top 2^{-1} \Omega^{-1} \tilde{g}_{+, (0,1)}(x_{2i})] - \mathbb{Z}_2(t) \right| \\ &\quad + \sup_{F \in \mathcal{F}} \sup_{t \in [t_0, T]} \left| \sqrt{n_t} \frac{1}{n_{1,t} n_{2,t}} \sum_{\mathbf{i} \in D_n(t)} K(\mathbf{z}_i) \right| = J_{n,1} + J_{n,2} + J_{n,3}, \end{aligned}$$

say, where $\mathbb{Z}_1(\cdot)$ and $\mathbb{Z}_2(\cdot)$ denote Gaussian processes. In (i) we proved that

$$\frac{\sqrt{n_t}}{n_{1,t} n_{2,t}} \sum_{\mathbf{i} \in D_n(t)} K(\mathbf{z}_i) \xrightarrow{P} 0.$$

In fact, this is true uniformly in t and F , thus $J_{n,3} \xrightarrow{P} 0$.

Use notations in the proof of [Theorem 1\(iii\)](#),

$$\begin{aligned} J_{n,1} &= \sqrt{n_1} \frac{1}{n_1} \sum_{i=1}^{n_1} \{ \tilde{h}_{(1,0)}(y_{1i}) - A^\top 2^{-1} \Omega^{-1} \tilde{g}_{+, (1,0)}(x_{1i}) \} v_{1,i} - \mathbb{Z}_1(t) \\ &= \sqrt{n_1} \frac{1}{n_1} \sum_{i=1}^{n_1} r(y_{1i}, x_{1,i}) - \mathbb{Z}_1(t) = \{ \mathbb{V}_{n,1}(t) - \mathbb{Z}_1(t) \}. \end{aligned}$$

Note that $r(\cdot, \cdot)$ implicitly depends on (F, t) , and on $n_{1,t}$ through $v_{1,i}$. For fixed F , let $\mathcal{B}_0 = \{r(\cdot, \cdot) : t \in [t_0, T]\}$, we are to prove \mathcal{B}_0 is a Donsker class uniformly on \mathcal{F} , but the dependence on $n_{1,t}$ is not easy to deal with. As in the proof of [Theorem 1\(iii\)](#), instead we define $\mathcal{B} = \{r(\cdot, \cdot) : t \in [t_0, T], \text{ with } v_{1,i} \text{ replaced by all functions } u_1(t), \text{ with } 0 \leq u_1(t) \leq 1, t \in [t_0, T]\}$. Then $\mathcal{B}_0 \subset \mathcal{B}$, and we only need to prove that \mathcal{B} is a uniform Donsker class over \mathcal{F} . It is apparent that \mathcal{B} has a bounded envelope $R(\cdot, \cdot) = 1 + |A^\top 2^{-1} \Omega^{-1} \tilde{g}_{+, (1,0)}(\cdot)|$, and it is easy to see that

$$\lim_{M \rightarrow \infty} \sup_{F \in \mathcal{F}} \sup_{t \in [t_0, T]} E_F[R^2(Y, X) \mathbf{1}\{R(Y, X) > M\}] \rightarrow 0.$$

For each $\epsilon > 0$, let $N_{[]}(\epsilon, \mathcal{B}, L_2(F))$ be the ϵ -bracketing number for the class \mathcal{B} . It can be checked that $N_{[]}(\epsilon \|R\|_{F,2}, \mathcal{B}, L_2(F)) = O(1/\epsilon)$ for all F , and so

$$\int_0^1 \sup_{F \in \mathcal{F}} \sqrt{\ln N_{[]}(\epsilon \|R\|_{F,2}, \mathcal{B}, L_2(F))} d\epsilon = O(1) \int_0^1 \sqrt{-\ln \epsilon} d\epsilon = O(1) \int_0^\infty \sqrt{y} e^{-y} dy < \infty.$$

Thus by Theorem 2.8.4 in VW (p. 172), \mathcal{B} is a uniform Donsker class over \mathcal{F} , i.e.,

$$\mathbb{V}_{n,1}(\cdot) \rightsquigarrow \mathbb{Z}_1(\cdot) \quad \text{in } \ell^\infty([t_0, T]), \text{ uniformly over } \mathcal{F}$$

where \mathbb{Z}_1 is a mean zero Gaussian process on $[t_0, T]$, and $\ell^\infty([t_0, T])$ is the space of all bounded real functions f on $[t_0, T]$ equipped with the sup norm $\|f\| = \sup_{t \in [t_0, T]} |f(t)|$. Note that the original condition of Theorem 2.8.4 in VW is

$$\int_0^\infty \sqrt{\ln N_{[]}(\epsilon \|R\|_{F,2}, \mathcal{B}, L_2(F))} d\epsilon < \infty.$$

Since for large ϵ , $N_{[]}(\epsilon \|R\|_{F,2}, \mathcal{B}, L_2(F)) = 1$, or $\ln N_{[]}(\epsilon \|R\|_{F,2}, \mathcal{B}, L_2(F)) = 0$, so the integral on $(0, \infty)$ is finite if and only if the integral on $(0, 1)$ is finite.

Lastly, we compute the covariance function $R(s, t)$. Using the results in (i), it can be seen that

$$\begin{aligned} R(s, t) &= \sum_{j=1}^2 \frac{(n_{1,s} n_{1,t})^{1/4} (n_{2,s} n_{2,t})^{1/4}}{n_{j,s} n_{j,t}} (n_{j,s} \wedge n_{j,t}) \tau_j^2 \\ &= c^{1/4} (s) c^{1/4} (t) b_1(s \wedge t, s \vee t) + c^{1/4} (s) c^{1/4} (t) b^{1/2}(s \wedge t, s \vee t) \\ &= c^{1/4} (s \wedge t) b^{1/2}(s \wedge t, s \vee t) \sum_{j=1}^2 c^{5/4-j} (s \vee t) \tau_j^2. \end{aligned}$$

Part (iii): Let $s = t/(T - t_0)$, $W_s = \sqrt{s} \sigma^{-1} \{(T - t_0)s\} \mathbb{Z}_1\{(T - t_0)s\}$. Then W_s is a mean zero Gaussian process on $[0, 1]$ with independent increments and covariance function given, for all $s, t \in [0, 1]$, by $s \wedge t$. From part (ii), under \mathcal{H}'_0 we have

$$\sup_{t \in [t_0, T]} \sqrt{n_t} \sqrt{t/(T - t_0)} \sigma^{-1}(t) U_n(t) = \sup_{s \in [0, 1]} \sqrt{n_{(T-t_0)s}} \sqrt{s} \sigma^{-1}\{(T - t_0)s\} U_n\{(T - t_0)s\} \rightsquigarrow \sup_{s \in [0, 1]} W_s$$

and by (10.17) in Billingsley [1] (p. 72), $\sup_{s \in [0, 1]} W_s = W$ with distribution function G as specified in the statement of the theorem.

Similarly, $\inf_{t \in [t_0, T]} \sqrt{n_t} \sqrt{t/(T - t_0)} \sigma^{-1}(t) U_n(t) \rightsquigarrow \inf_{s \in [0, 1]} W_s$. Note that $-W_s$ and W_s have the same distribution, so for $w \leq 0$,

$$\begin{aligned} \Pr(\inf_{s \in [0, 1]} W_s \leq w) &= 1 - \Pr(\inf_{s \in [0, 1]} W_s > w) \\ &= 1 - \Pr(\sup_{s \in [0, 1]} -W_s < -w) = 1 - \Pr(\sup_{s \in [0, 1]} W_s < -w) \\ &= 1 - G(-w). \end{aligned}$$

This completes the proof of [Theorem 2](#). \square

Proof of Corollary 1. The relationships $E\{D(n_{1,t}, n_{2,t})\} = \delta(t)I(n_{1,t}, n_{2,t}) + o(1)$ and $\text{var}\{D(n_{1,t}, n_{2,t})\} = I(n_{1,t}, n_{2,t}) + o(1)$ are direct consequence of Theorem 2. In the proof of Theorem 2(i), we have

$$\begin{aligned}\sqrt{n_t}\{U_n(t) - \bar{\theta}\} &= \frac{c^{1/4}(t)}{\sqrt{n_{1,t}}} \sum_{i=1}^{n_{1,t}} \{\tilde{h}_{(1,0)}(y_{1i}) - A^\top 2\Omega^{-1} \tilde{\tilde{g}}_{(1,0)}(x_{1i})\} \\ &\quad + \frac{c^{-1/4}(t)}{\sqrt{n_{2,t}}} \sum_{i=1}^{n_{2,t}} \{\tilde{h}_{(0,1)}(y_{2i}) - A^\top 2\Omega^{-1} \tilde{\tilde{g}}_{(0,1)}(x_{2i})\} + o_P(1)\end{aligned}$$

or

$$\begin{aligned}D(n_{1,t}, n_{2,t}) - I(n_{1,t}, n_{2,t})\delta(t) &= \sigma^{-1}(t) \left[c^{1/2}(t) \sum_{i=1}^{n_{1,t}} \{\tilde{h}_{(1,0)}(y_{1i}) - A^\top 2\Omega^{-1} \tilde{\tilde{g}}_{(1,0)}(x_{1i})\} \right. \\ &\quad \left. + c^{-1/2}(t) \sum_{i=1}^{n_{2,t}} \{\tilde{h}_{(0,1)}(y_{2i}) - A^\top 2\Omega^{-1} \tilde{\tilde{g}}_{(0,1)}(x_{2i})\} \right] + o_P(1).\end{aligned}$$

The above and the independence of the y_{1i} 's and y_{2i} 's together yield

$$\text{cov}\{D(n_{1,s}, n_{2,s}), D(n_{1,t}, n_{2,t})\} = \text{var}\{D(n_{1,s}, n_{2,s})\} + o(1).$$

Thus the proof of Corollary 1 is complete. \square

Proof of Theorem 3. As in the proofs of the previous theorems, with g_+ replaced by G , since $r_0 = \min\{\text{rank}(g_1), \dots, \text{rank}(g_d), \text{rank}(h)\} = 1$, by Lemma (ii) we have

$$w_i = \frac{1}{n_{1,t}n_{2,t}} \left[1 - G^\top(\mathbf{z}_i|\bar{\theta})\Lambda^{-1} \frac{1}{n_{1,t}n_{2,t}} \sum_{j \in D_n(t)} G(\mathbf{z}_j|\bar{\theta}) + \{1_{d+1}^\top G(\mathbf{z}_i|\bar{\theta}) + \|G(\mathbf{z}_i|\bar{\theta})\|^2\} O_P(n_t^{-1}) \right],$$

and by two-sample U-statistics theory (as in Theorem 4.5.1, KB, p. 151),

$$\sqrt{n_t} \frac{1}{n_{1,t}n_{2,t}} \sum_{j \in D_n(t)} G(\mathbf{z}_j|\bar{\theta}) \rightsquigarrow \mathcal{N}(0, \Lambda_1), \quad \frac{1}{n_{1,t}n_{2,t}} \sum_{i \in D_n(t)} G(\mathbf{z}_i|\bar{\theta})G^\top(\mathbf{z}_i|\bar{\theta}) = \Lambda + O_P(n_t^{-1/2}).$$

Furthermore,

$$\begin{aligned}\lambda &= \Lambda^{-1} \frac{1}{n_{1,t}n_{2,t}} \sum_{j \in D_n(t)} G(\mathbf{z}_j|\bar{\theta}) + O_P(n_t^{-1}) = O_P(n_t^{-1/2}), \\ \frac{1}{n_{1,t}n_{2,t}} \sum_{i \in D_n(t)} \|G(\mathbf{z}_i|\bar{\theta})\|^2 &\xrightarrow{\text{a.s.}} E_F\{\|G(\mathbf{Z}|\bar{\theta})\|^2\} < \infty,\end{aligned}$$

and $\max_i |\lambda' G(\mathbf{z}_i|\bar{\theta})| = O_P(n_t^{-1/2} n_t^{1/\alpha}) \xrightarrow{P} 0$ since $1/\alpha < r_0/2 = 1/2$, so

$$-\frac{n_t}{n_{1,t}n_{2,t}} \ln R_G(\bar{\theta}) = \frac{n_t}{n_{1,t}n_{2,t}} \sum_{i \in D_n(t)} \ln\{1 + \lambda^\top G(\mathbf{z}_i|\bar{\theta})\}.$$

The latter can be rewritten successively as

$$\frac{n_t}{n_{1,t}n_{2,t}} \sum_{i \in D_n(t)} \left\{ \lambda^\top G(\mathbf{z}_i|\bar{\theta}) - \frac{1}{2} \lambda^\top G(\mathbf{z}_i|\bar{\theta}) G(\mathbf{z}_i|\bar{\theta}) \lambda + o_P(n^{-1}) \|G(\mathbf{z}_i|\bar{\theta})\|^2 \right\},$$

and

$$\begin{aligned}&n_t \frac{1}{n_{1,t}n_{2,t}} \sum_{j \in D_n(t)} G^\top(\mathbf{z}_j|\bar{\theta}) \Lambda^{-1} \frac{1}{n_{1,t}n_{2,t}} \sum_{i \in D_n(t)} G(\mathbf{z}_i|\bar{\theta}) - \frac{n_t}{2n_{1,t}n_{2,t}} \sum_{i \in D_n(t)} \left\{ \frac{1}{n_{1,t}n_{2,t}} \sum_{j \in D_n(t)} G^\top(\mathbf{z}_j|\bar{\theta}) \Lambda^{-1} \right\} \\ &\quad \times G(\mathbf{z}_i|\bar{\theta}) G^\top(\mathbf{z}_i|\bar{\theta}) \left\{ \Lambda^{-1} \frac{1}{n_{1,t}n_{2,t}} \sum_{j \in D_n(t)} G(\mathbf{z}_j|\bar{\theta}) \right\} + o_P(n_t^{-1}) \frac{n_t}{n_{1,t}n_{2,t}} \sum_{i \in D_n(t)} \|G(\mathbf{z}_i|\bar{\theta})\|^2,\end{aligned}$$

which can be further written as

$$\begin{aligned}&n_t \frac{1}{n_{1,t}n_{2,t}} \sum_{j \in D_n(t)} G^\top(\mathbf{z}_j|\bar{\theta}) \Lambda^{-1} \frac{1}{n_{1,t}n_{2,t}} \sum_{i \in D_n(t)} G(\mathbf{z}_i|\bar{\theta}) - n_t \left\{ \frac{1}{2n_{1,t}n_{2,t}} \sum_{j \in D_n(t)} G^\top(\mathbf{z}_j|\bar{\theta}) \Lambda^{-1} \right\} \\ &\quad \times \frac{1}{n_{1,t}n_{2,t}} \sum_{i \in D_n(t)} G(\mathbf{z}_i|\bar{\theta}) G^\top(\mathbf{z}_i|\bar{\theta}) \left\{ \Lambda^{-1} \frac{1}{n_{1,t}n_{2,t}} \sum_{j \in D_n(t)} G(\mathbf{z}_j|\bar{\theta}) \right\} + o_P(1)\end{aligned}$$

and

$$n_t \frac{1}{n_{1,t} n_{2,t}} \sum_{j \in D_n(t)} G^\top(\mathbf{z}_j | \bar{\theta}) \Lambda^{-1} \frac{1}{n_{1,t} n_{2,t}} \sum_{i \in D_n(t)} G(\mathbf{z}_i | \bar{\theta}) - n_t \frac{1}{2n_{1,t} n_{2,t}} \sum_{j \in D_n(t)} G(\mathbf{z}_j | \bar{\theta}) \Lambda^{-1} \\ \times \left\{ \Lambda + O_p(n_t^{-1/2}) \right\} \Lambda^{-1} \frac{1}{n_{1,t} n_{2,t}} \sum_{i \in D_n(t)} G^\top(\mathbf{z}_i | \bar{\theta}) + o_p(1),$$

which finally reduces to

$$n_t \frac{1}{n_{1,t} n_{2,t}} \sum_{j \in D_n(t)} G^\top(\mathbf{z}_j | \bar{\theta}) \Lambda^{-1} \frac{1}{n_{1,t} n_{2,t}} \sum_{i \in D_n(t)} G(\mathbf{z}_i | \bar{\theta}) - O_p(n^{-1/2}) + o_p(1).$$

This completes the proof since

$$\sqrt{n_t} \Lambda_1^{-1/2} \frac{1}{n_{1,t} n_{2,t}} \sum_{i \in D_n(t)} G(\mathbf{z}_i | \bar{\theta}) \rightsquigarrow \mathcal{N}(\mathbf{0}, I_{r+1}).$$

Finally, the second statement is a special case. \square

References

- [1] P. Billingsley, *Convergence in Probability Measures*, Wiley, New York, 1968.
- [2] W. Hoeffding, A class of statistics with asymptotically normal distribution, *Ann. Math. Statist.* 19 (1948) 293–325.
- [3] W. Hoeffding, The strong law of large numbers for U-statistics, *Inst. Statist. Mimeo. Ser.* 302 (1961) 1–10.
- [4] P. Huang, M.T. Tan, Multistage nonparametric tests for treatment comparisons in clinical trials with multiple primary endpoints, *Stat. Interface* 9 (2016) 343–354.
- [5] P. Huang, B.C. Tilley, R.F. Woolson, S. Lipsitz, Adjusting O'Brien's test to control type I error for the generalized nonparametric Behrens–Fisher problem, *Biometrics* 61 (2005) 532–539.
- [6] P. Huang, R.F. Woolson, P.C. O'Brien, A rank-based sample size method for multiple outcomes in clinical trials, *Stat. Med.* 27 (2008) 3084–3104.
- [7] C. Jennison, B.W. Turnbull, *Group Sequential Methods with Applications to Clinical Trials*, CRC Press, Boca Raton, FL, 2000.
- [8] B.Y. Jing, J. Yuan, W. Zhou, Empirical likelihood for U-statistics, *J. Amer. Statist. Assoc.* 104 (2009) 1224–1232.
- [9] G.G. Koch, I.A. Amara, G.W. Davis, D.B. Gillings, A review of some statistical methods for covariance analysis of categorical data, *Biometrics* 38 (1982) 563–595.
- [10] G.G. Koch, C.M. Tangen, J.-W. Jung, I.A. Amara, Issues for covariate analysis of dichotomous and ordered categorical data from randomized clinical trials and non-parametric strategies for addressing them, *Stat. Med.* 17 (1988) 1863–1892.
- [11] V.S. Koroljuk, Yu.V. Borovskich, *Theory of U-Statistics*, Kluwer Academic Publishers, The Netherlands, 1994.
- [12] M.R. Kosorock, S. Yuanjun, D.L. DeMets, Design and analysis of group sequential clinical trials with multiple primary endpoints, *Biometrics* 60 (2004) 134–145.
- [13] H. Lin, Y. Li, M.T. Tan, Estimating a unitary effect summary based on combined survival and quantitative outcomes, *Comput. Statist. Data Anal.* 66 (2013) 129–139.
- [14] A. Liu, Q. Li, C. Liu, K. Yu, K.F. Yu, A rank-based test for comparison of multidimensional outcomes, *J. Amer. Statist. Assoc.* 105 (2010) 578–587.
- [15] X. Lu, A.A. Tsiatis, Improving the efficiency of the log-rank test using auxiliary covariates, *Biometrika* 95 (2008) 679–694.
- [16] P.C. O'Brien, Procedures for comparison samples with multiple end-points, *Biometrics* 40 (1984) 1079–1087.
- [17] P.C. O'Brien, T.R. Fleming, A multiple testing procedure for clinical trials, *Biometrics* 35 (1979) 549–556.
- [18] A.B. Owen, Empirical likelihood confidence regions, *Ann. Statist.* 18 (1990) 90–120.
- [19] A.B. Owen, *Empirical Likelihood*, Chapman & Hall, London, 2001.
- [20] J. Qin, J.F. Lawless, Empirical likelihood and general estimating equations, *Ann. Statist.* 22 (1994) 300–325.
- [21] A. Schacht, K. Bogaerts, E. Bluhmki, E. Lesaffre, A new nonparametric approach for baseline covariate adjustment for two-group comparative studies, *Biometrics* 64 (2008) 1110–1116.
- [22] R. Serfling, *Approximation Theorems of Mathematical Statistics*, Wiley, New York, 1980.
- [23] M. Tan, X. Xiong, M.H. Kutner, Clinical trial designs based on sequential conditional probability ratio tests and reverse stochastic curtailing, *Biometrics* 54 (1998) 682–695.
- [24] The Parkinson Study Group, Effects of tocopherol and deprenyl on the progression of disability in early Parkinson disease, *N. Engl. J. Med.* 328 (1993) 1761–1763.
- [25] A.A. Tsiatis, M. Davidian, M. Zhang, X. Lu, Covariate adjustment for two-sample treatment comparisons in randomized clinical trials: A principled yet flexible approach, *Stat. Med.* 27 (2008) 4658–4677.
- [26] A. van der Vaart, J. Wellner, *Weak Convergence and Empirical Processes: With Applications to Statistics*, Springer, New York, 1996.
- [27] A.T.A. Wood, K.A. Do, N.M. Broom, Sequential linearization of empirical likelihood constraints with application to U-statistics, *J. Comput. Graph. Statist.* 5 (1996) 365–385.
- [28] X. Xiong, M. Tan, J. Boyett, Sequential conditional probability ratio tests for normalized test statistic on information time, *Biometrics* 59 (2003) 624–631.
- [29] A. Yuan, W. He, B. Wang, G. Qin, U-statistic with side information, *J. Multivariate Anal.* 111 (2012) 20–38.
- [30] M. Zhang, A.A. Tsiatis, M. Davidian, Improving efficiency of inferences in randomized clinical trials using auxiliary covariates, *Biometrics* 64 (2008) 707–715.