

# ROC Analysis for Phase II Group Sequential Basket Clinical Trial

Sirao Wang<sup>1</sup>, Ao Yuan<sup>1,2,\*</sup>, Larry Tang<sup>2,3</sup>, Hong-Bin Fang<sup>1</sup>, Ming T. Tan<sup>1</sup> and Leighton Chan<sup>2</sup>

<sup>1</sup>Department of Biostatistics, Bioinformatics and Biomathematics, Georgetown University, Washington DC 20057, USA

<sup>2</sup>NIH Clinical Center, Rehabilitation Medicine Department, Bethesda, MD 20892, USA

<sup>3</sup>Department of Statistics, George Mason University, Fairfax, VA 22030, USA

**Abstract:** The basket trial is a recent development in the clinical trial practice. It conducts the test of the same treatment on several different related diseases in a single trial, and has the advantage of reduced cost and enhanced efficiency. A natural question is how to assess the performance of the group sequential basket trial against the classical group sequential trial? To our knowledge, a formal assessment hasn't been seen in the literature, and is the goal of this study. Specifically, we use the receiver operating characteristic curve to assess the performance of the mentioned two trials. We considered two cases, parametric and nonparametric settings. The former is efficient when the parametric model is correctly specified, but can be misleading if the model is incorrect; the latter is less efficient but is robust in that it cannot be wrong no matter what the true data generating model is. Simulation studies are conducted to evaluate the experiments, and it suggests that the group sequential basket trial generally outperforms the group sequential trial in either the parametric and nonparametric cases, and that the nonparametric method gives more accurate evaluation than the parametric one for moderate to large sample sizes.

**Keywords:** Basket trial, group sequential clinical trial, nonparametric ROC curve, parametric ROC curve, phase II clinical trial.

## 1. INTRODUCTION

The basket clinical trial design (For example, [1-4]) is a recent practice in the clinical trial field. Different from traditional clinical trials, which examines one treatment for one targeted disease, the basket design examines one treatment on several different (but often related) diseases in a single trial. By this way, it explores much more potential of the treatment, and save possible costs and time, if separate trials were conducted on the different diseases. Another motivation for this type of design is to examine a common response (such as a biomarker response) across multiple diseases (tumors). The number of patients with a putative biomarker within a single disease is small, which makes it difficult to enroll adequate number of patients in a conventional trial, and the basket trial which pool the responses from the same biomarker from all the patients with different diseases makes the trial possible. The basic assumption under the rationale for basket trial is that the fundamental classification of disease is the response, not disease type [5-9]. The disadvantage of this trial is that inactive responses from some disease patients may dilute the pooled signal and trigger failure of the entire trial. Thus this type of trial have been used primarily for exploratory settings [10].

The receiver operating characteristic (ROC) curve is common tool for evaluating the performance of test statistic. It is the plot of the true positive rate (TPR) (i.e. probability of identifying a diseased subject when the subject is truly diseased) versus false positive rate (FPR) (i.e. probability of identifying a diseased subject when the subject is not diseased). It is also widely used in biostatistics, medical diagnostic biomarkers, radiology, psychophysical and medical imaging research, military monitoring, and industrial quality control (Metz, 1978) [11]. The ROC curve indicates the trade-off between the TPR and FPR under different thresholds. It has many advantages and overcomes the limitation of using isolated measurements of TPR and FPR. The ROC curve is plotted by connecting all the points generated by possible thresholds [12]. It has wide applications in biomedical research works (for example, [13, 14]).

Here we consider the ROC analysis for basket phase II clinical trial. Two cases are considered, the parametric and nonparametric settings. The former is efficient when the parametric model is correctly specified, but can be misleading if the model is incorrect; the latter is less efficient but is robust in that it cannot be wrong no matter what the true data generating model is. Simulation studies are conducted to evaluate the experiments, and it suggests that the group sequential basket trial generally outperforms the group sequential trial in either the parametric and nonparametric cases, and that the nonparametric method gives more accurate evaluation than the parametric one for moderate to large sample sizes.

\*Address correspondence to this author at the Department of Biostatistics, Bioinformatics and Biomathematics, Georgetown University, Washington DC 20057, USA; Tel: 202-687-0766; E-mail: yuanao@hotmail.com

**2. THE METHOD**

Group sequential (or multi-stage) design is commonly used in phase II and III clinical trials to evaluate a new treatment against some existing one(s) [15-24]. In contrast to the none-sequential clinical trial, the group sequential trial has the feature that it can early stop the trial before the planned end, if extreme outcome is detected at some intermediate stage. Thus, group sequential basket design is natural in stage II clinical trial. A natural question is how to assess the performance of the group sequential basket trial against the classical group sequential trial? To our knowledge, a formal assessment hasn't been seen in the literature, and is the goal of this study. Specifically, we use the receiver operating characteristic curve, a commonly used tool for assessing the performance of test statistic, for our evaluations.

In this study we concentrate on continuous response. The trial has  $k$  stages, at stage  $l$ , there are independent responses  $x_{ij}$  for the  $i$ -th patient with disease type  $j$ ,  $j = 1, \dots, d; i = 1, \dots, n_{lj} (l = 1, \dots, k)$ . Note that for fixed  $i$ ,  $x_{ij}$  has at least one non-zero entry but not all of them non-zero, as each patient may not have all the diseases. For each fixed  $j$ ,  $n_{lj} > n_{rj}$  for  $r > l$ , and we assume that the non-zero  $x_{ij}$ 's are iid  $N(\mu_j, \sigma_j^2)$ . For phase II clinical trial, often the total number of patients  $n = \sum_{l=1}^k \sum_{j=1}^d n_{lj}$  is small (typically  $10 < n < 100$ ),  $2 \leq d \leq 10$  and  $2 \leq k \leq 10$ . The hypothetical population mean response is  $\mu = (\mu_1, \dots, \mu_d)'$ . We are interested in testing the null hypothesis

$$H_{0j}: \mu_j \leq \mu_{0j}, \text{ vs } H_{1j}: \mu_j > \mu_{1j} = \mu_{0j} + \delta_j, (j = 1, \dots, d) \tag{1}$$

where  $\mu_0 = (\mu_{01}, \dots, \mu_{0d})'$  is the given vector of threshold values for the responses to be effective, and  $\delta$  is the vector of pre-specified meaningful differences for each of the diseases. Although the observations  $x_{ij}$ 's are independent, but the diseases themselves are dependent via the shared common factor(s), for example, the common marker(s) which brought the patients to the trial. Denote  $x_i = (x_{i1}, \dots, x_{id})'$ , and  $\mathbf{X} = (X_1, \dots, X_d)$  be an iid copy of it but without zero entry. We use frailty to model the dependence among the disease responses  $\mathbf{X}$ . Let  $C$  be the shared common factor of the diseases, we assume that conditioning on  $C$ , the responses form the diseases are independent, thus the joint law of  $\mathbf{X}$  is given by  $P(\mathbf{X}) = \int P(\mathbf{X}|C = c)P(c)dc = \int \prod_{j=1}^d P(X_j|C = c)P(c)dc$ . In particular, we assume  $P(X_j|C = c) = N(\mu_0 + c\eta_j, \sigma_j^2)$ ,  $\eta_j$  is a parameter to be estimated, and  $C \sim N(\theta, \tau^2)$ , where  $(\theta, \tau^2)$  can either be obtained from prior studies, or estimate from the current data. Assume  $b - \eta_j^2 > 0 (j = 1, \dots, d)$ , then

$$\mathbf{X} \sim N(\boldsymbol{\beta}, \Omega), \tag{2}$$

where,  $\boldsymbol{\beta}, \Omega, b$  and related notations are given in the Appendix. Thus we can assume  $Cov(x_i) = \Omega$ , the Pearson covariance matrix.

Now let  $\Omega$  be the given covariance matrix. Typically the test statistic  $S_l = (S_{l1}, \dots, S_{ld})'$  are of the form

$$S_l = \Omega^{-\frac{1}{2}} \left( \frac{\mathbf{1}}{\sqrt{n_{k1}}} \sum_{i=1}^{n_{l1}} x_{i1}, \dots, \frac{\mathbf{1}}{\sqrt{n_{kd}}} \sum_{i=1}^{n_{ld}} x_{id} \right)', (l = 1, \dots, k). \tag{3}$$

Thus, under the right boundary of  $H_0$ ,  $S_l \sim N(\boldsymbol{\mu}_{lk}, \mathbf{D}_{lk})$ , where  $\boldsymbol{\mu}_{lk} = \Omega^{-1/2} (\mu_{01}n_{l1}/\sqrt{n_{k1}}, \dots, \mu_{0d}n_{ld}/\sqrt{n_{kd}})'$  and  $\mathbf{D}_{lk} = \text{diag}\{n_{l1}/n_{k1}, \dots, n_{ld}/n_{kd}\}$ , a  $d$ -dimensional diagonal matrix.

In practice, it is also of interest to test the effect of the treatment on each of the disease types, which can be formulated as  $H_{0j}: \mu_j \leq \mu_{0j}$  vs  $H_{aj}: \mu_j > \mu_{1j} = \mu_{0j} + \delta_j, (j = 1, \dots, d)$ . If  $H_{0j}$  is rejected, we conclude that the treatment is likely to be effective for this disease.

To test  $H_{0j}$  vs  $H_{aj}$ , a simple way is just use the statistic  $n_{kj}^{-1/2} \sum_{i=1}^{n_{lj}} x_{ij}$  at the  $l$ -th interim stage ( $l = 1, \dots, k$ ), for disease type  $j (j = 1, \dots, d)$ . However, this is the classical trial, not the basket trial. In the latter trial, we want to use the information across all the diseases to perform each single hypothesis.

For this, at the  $l$ -th stage, we want to test  $H_{0j}$  vs  $H_{aj} (j = 1, \dots, d)$  to see which disease type(s) are significant, which are not. Let  $n_l = \sum_{j=1}^d n_{lj} (l = 1, \dots, k)$ . To borrow information from all the disease types, let  $\bar{S}_l = \sum_{j=1}^d \frac{n_{lj}}{n_l} S_{lj}$  and  $\bar{\mu}_l = \sum_{j=1}^d \frac{n_{lj}}{n_l} \mu_{lj}$ , we use the conditional statistic

$$S_{l,j} | \bar{S}_l \sim N(\tilde{\mu}_j, \tilde{\sigma}_j^2), \tag{4}$$

where  $\tilde{\mu}_j = \mu_{lj} + n_{lj}^2 / (n_{kj}n_l) D_l^{-1} (\bar{S}_l - \bar{\mu}_l)$ ,  $D_l = \sum_{j=1}^d n_{lj}^3 / (n_{kj}n_l^2)$ , and  $\tilde{\sigma}_j^2 = n_{lj} / n_{kj} - n_{lj}^4 / (n_{kj}n_l) D_l^{-1}$ . The level  $\alpha$  cut-off point is obtained accordingly, which depends on  $\bar{S}_l$  and is random.

If  $H_{0j}$  is rejected at stage  $l$ , then data on the  $j$ -th disease will be removed, and the trial moves on based on the remaining data.

**2.1. Brief Review of ROC and AUC Analysis**

The receiver operating characteristic (ROC) curve and the area under the curve (AUC) are common tools for assessing the performance of a test statistic. Here we derive the ROC and AUC for our case. We first give a brief review of these concepts. The ROC and AUC are used to characterize the relationship between the true positive rate (TPR) and the false positive rate (FPR) of a test statistic  $T$ . TPR is the probability of identifying an abnormal location when the result of this location is truly

abnormal, and FPR is the probability of identifying an abnormal location when the result of this location is truly normal. Let  $D$  denote the disease status,  $D = 1$  means truly diseased,  $D = 0$  for no disease. Let  $F$  be the distribution function of  $T$  under the null hypothesis of no disease, and  $G$  be that of  $T$  under the alternative, given a threshold value  $t$ ,  $TPR(t) = P(T > t|D = 1) = 1 - G(t)$  and  $FPR(t) = P(T > t|D = 0) = 1 - F(t)$ , and the ROC curve is defined as

$$ROC(u) = 1 - G(F^{-1}(1 - u)), 0 < u < 1. \tag{5}$$

Typically,  $T$  is constructed such that large value of  $T$  corresponds to evidence in favor of the alternative, so  $P(T > t|D = 1) = 1 - G(t) \geq P(T > t|D = 0) = 1 - F(t)$ , or  $G(\cdot) \leq F(\cdot)$ ,  $F^{-1}(\cdot) \leq G^{-1}(\cdot)$ , and as a result  $ROC(u) \geq u (0 < u < 1)$ , with " $\geq$ " only if  $G(\cdot) = F(\cdot)$ . Let  $X = (T|D = 0)$  and  $Y = (T|D = 1)$ . The AUC is defined as the mean value of the ROC curve,

$$AUC = \int_0^1 ROC(u) du = P(Y > X). \tag{6}$$

The last equality follows as  $P(Y > X) = \int [1 - G(x)]dF(x) = \int_0^1 [1 - G(F^{-1}(v))]dv = \int_0^1 [1 - G(F^{-1}(1 - u))]du$ . It is apparent that  $0.5 \leq AUC \leq 1$ . If we have two treatments, then we have two ROC curves, and their performances are evaluated by comparing their ROC curves. Large AUC corresponds to better test (treatments).

### 2.2. Parametric ROC

In many clinical trial studies, parametric models are used for testing the hypothesis and ROC analysis. In particular, the normal model is commonly used:  $F \sim N(\mu_1, \sigma_1^2)$  and  $G \sim N(\mu_2, \sigma_2^2)$ . Let  $\Phi(\cdot)$  be the distribution function of  $N(0,1)$  and  $\Phi^{-1}(\cdot)$  its inverse or quantile function. Write  $F^{-1}(1 - u) = t$ , or

$$1 - u = F(t) = \Phi((t - \mu_1)/\sigma_1), \text{ and get } F^{-1}(1 - u) = t = \mu_1 + \sigma_1 \Phi^{-1}(1 - u). \text{ In this case,}$$

$$\begin{aligned} ROC(u) &= 1 - \Phi\left(\frac{F^{-1}(1 - u) - \mu_2}{\sigma_2}\right) \\ &= 1 - \Phi\left(\frac{\mu_1 - \mu_2}{\sigma_2} + \frac{\sigma_1}{\sigma_2} \Phi^{-1}(1 - u)\right) \\ &= \Phi\left(\frac{\mu_2 - \mu_1}{\sigma_2} - \frac{\sigma_1}{\sigma_2} \Phi^{-1}(1 - u)\right) = \Phi\left(\frac{\mu_2 - \mu_1}{\sigma_2} + \frac{\sigma_1}{\sigma_2} \Phi^{-1}(u)\right), u \in [0,1]. \end{aligned} \tag{7}$$

The last equality above holds because for a standard normal random variable  $Z$ ,  $1 - \Phi(v) = P(Z > v) = P(Z < -v) = \Phi(-v)$ ; and set  $\Phi^{-1}(1 - u) = t$ , then  $u = 1 - \Phi(t) = P(Z > t) = P(Z < -t) = \Phi(-t)$ , which gives  $\Phi^{-1}(1 - u) = t = -\Phi^{-1}(u)$ .

Now come to our problem. Suppose at stage  $l$ , we observe  $S_l^0 = (S_{l1}^0, \dots, S_{ld}^0)'$  from the standard treatment, and  $S_l^1 = (S_{l1}^1, \dots, S_{ld}^1)'$  from the new treatment ( $l = 1, \dots, k$ ).

$$S_l^0 = \Omega_0^{-1/2} \left( \sum_{i=1}^{n_{l1}^0} x_{i1} / \sqrt{n_{k1}^0}, \dots, \sum_{i=1}^{n_{ld}^0} x_{id} / \sqrt{n_{kd}^0} \right)' \sim N(\mu_{lk}^0, D_{lk}^0),$$

with  $\mu_{lk}^0 = \Omega_0^{-1/2} (\mu_{01} n_{l1}^0 / \sqrt{n_{k1}^0}, \dots, \mu_{0d} n_{ld}^0 / \sqrt{n_{kd}^0})'$  and  $D_{lk}^0 = \text{diag}\{n_{l1}^0/n_{k1}^0, \dots, n_{ld}^0/n_{kd}^0\}$ .

$$S_l^1 = \Omega_1^{-1/2} \left( \sum_{i=1}^{n_{l1}^1} y_{i1} / \sqrt{n_{k1}^1}, \dots, \sum_{i=1}^{n_{ld}^1} y_{id} / \sqrt{n_{kd}^1} \right)' \sim N(\mu_{lk}^1, D_{lk}^1)$$

with  $\mu_{lk}^1 = \Omega_1^{-1/2} (\mu_{11} n_{l1}^1 / \sqrt{n_{k1}^1}, \dots, \mu_{1d} n_{ld}^1 / \sqrt{n_{kd}^1})'$  and  $D_{lk}^1 = \text{diag}\{n_{l1}^1/n_{k1}^1, \dots, n_{ld}^1/n_{kd}^1\}$ .

We want test  $H_{0,j}: \mu_{1j} \leq \mu_{0j}$  vs  $H_{a,j}: \mu_{1j} > \mu_{0j} = \mu_{0j} + \delta_j$ , ( $j = 1, \dots, d$ ). Let  $n_l^0 = \sum_{j=1}^d n_{lj}^0$  and  $n_l^1 = \sum_{j=1}^d n_{lj}^1$ , ( $l = 1, \dots, k$ ). To borrow information from all the disease types, let  $\bar{S}_l^0 = \sum_{j=1}^d \frac{n_{lj}^0}{n_l^0} S_{lj}^0$  and  $\bar{\mu}_l^0 = \sum_{j=1}^d \frac{n_{lj}^0}{n_l^0} \mu_{lj}^0$ , we use the conditional statistic

$$S_{l,j}^0 | \bar{S}_l^0 \sim N(\bar{\mu}_j^0, (\bar{\sigma}_{lj}^0)^2),$$

where  $\bar{\mu}_j^0 = \mu_{lj}^0 + (n_{lj}^0)^2 / (n_{kj}^0 n_l^0) D_{l0}^{-1} (\bar{S}_l^0 - \bar{\mu}_l^0)$ ,  $D_{l0} = \sum_{j=1}^d (n_{lj}^0)^3 / (n_{kj}^0 (n_l^0)^2)$ , and  $(\bar{\sigma}_{lj}^0)^2 = n_{lj}^0 / n_{kj}^0 - (n_{lj}^0)^4 / (n_{kj}^0 n_l^0)^2 D_{l0}^{-1}$ . The level  $\alpha$  cut-off point is obtained accordingly, which depends on  $\bar{S}_l^0$  and is random. Similarly, let  $n_l^1 = \sum_{j=1}^d n_{lj}^1$ ,  $\bar{S}_l^1 = \sum_{j=1}^d \frac{n_{lj}^1}{n_l^1} S_{lj}^1$  and  $\bar{\mu}_l^1 = \sum_{j=1}^d \frac{n_{lj}^1}{n_l^1} \mu_{lj}^1$ ,

$$S_{l,j}^1 | \bar{S}_l^1 \sim N(\bar{\mu}_j^1, (\bar{\sigma}_{lj}^1)^2),$$

where  $\bar{\mu}_j^1 = \mu_{lj}^1 + (n_{lj}^1)^2 / (n_{kj}^1 n_l^1) D_{l1}^{-1} (\bar{S}_l^1 - \bar{\mu}_l^1)$ ,  $D_{l1} = \sum_{j=1}^d (n_{lj}^1)^3 / (n_{kj}^1 (n_l^1)^2)$ , and  $(\bar{\sigma}_{lj}^1)^2 = n_{lj}^1 / n_{kj}^1 - (n_{lj}^1)^4 / (n_{kj}^1 n_l^1)^2 D_{l1}^{-1}$ .

The ROC curve of the basket trial for the  $j$ -th disease at the  $l$ -th interim trial is

$$\begin{aligned} \tilde{ROC}_{l,j}(u) &= \Phi\left(\frac{\bar{\mu}_{lj}^1 - \bar{\mu}_{lj}^0}{\bar{\sigma}_{lj}^1} + \frac{\bar{\sigma}_{lj}^0}{\bar{\sigma}_{lj}^1} \Phi^{-1}(u)\right), u \in [0,1] \\ (j = 1, \dots, d; l = 1, \dots, k). \end{aligned} \tag{8}$$

In contrast, the classical trial uses data  $S_{l,j}^1$  and  $S_{l,j}^0$ , separately for each  $j$  without conditioning on  $\bar{S}_l^1$  and  $\bar{S}_l^0$ , and the corresponding ROC curve for the  $j$ -th disease at the  $l$ -th interim trial is

$$ROC_{l,j}(u) = \Phi\left(\frac{\mu_{lj}^1 - \mu_{lj}^0}{\sigma_{lj}^1} + \frac{\sigma_{lj}^0}{\sigma_{lj}^1} \Phi^{-1}(u)\right), u \in [0,1] (j = 1, \dots, d; l = 1, \dots, k),$$

with  $(\sigma_{lj}^1)^2 = n_{lj}^1 / n_{kj}^1$  and  $(\sigma_{lj}^0)^2 = n_{lj}^0 / n_{kj}^0$ . Since  $\bar{\sigma}_{lj}^1 \leq \sigma_{lj}^1$  and  $\bar{\sigma}_{lj}^0 \leq \sigma_{lj}^0$ , ROC curve of the basket group

sequential trial is expected to have better performance than that of the classical group sequential trial.

### 2.3. Nonparametric ROC

When a specific parametric model is not justified for the observed data, the nonparametric method is preferred for robustness. In this case, the ROC curve will be constructed using the empirical distribution functions. However, different from the classical clinical trial, for the case of basket trial, to construct the empirical distribution for each disease, we need to borrow information from data on other diseases. To incorporate such shared side information, we use the method of empirical likelihood.

We consider the set-up for empirical likelihood (EL) as in Qin and Lawless (1994) [25]. We first give a general description of the method. Suppose the side information can be incorporated into the EL through a  $d$ -dimensional known function  $g(x) = (g_1(x), \dots, g_d(x))'$  via the relationship

$$E[g(X_1)] = \mathbf{0},$$

where  $E[\cdot]$  denotes the expectation with respect to the data distribution function  $F$ . Since the empirical likelihood is a nonparametric maximum likelihood estimate of the distribution function, under some possible constraint(s), it is a step function with jumps at the data points, so we can set  $w_i = F(X_i) (i = 1, \dots, n)$ . The EL is defined as

$$L(F) = \prod_{i=1}^n w_i,$$

where the  $w_i$ s are the nonparametric maximum likelihood estimated empirical masses assigned to the observation  $X_i$ s. With the side information constraints, the EL is

$$\max_w \prod_{i=1}^n w_i \text{ subject to } \sum_{i=1}^n w_i = 1 \text{ and } \sum_{i=1}^n w_i g(X_i) = \mathbf{0}.$$

Let  $t = (t_1, \dots, t_d)'$  be the Lagrange multipliers corresponding to the constraint of  $g(\cdot)$ , and as in Owen (1990) [26], we get

$$w_i = \frac{1}{n} \frac{1}{1 + t'g(X_i)},$$

where  $t_j = t_j(X_1, \dots, X_n) (j = 1, \dots, d)$  are determined by

$$\sum_{i=1}^n \frac{g(X_i)}{1 + t'g(X_i)} = \mathbf{0}.$$

Then the weighted empirical distribution function with the side information incorporated is

$$\tilde{F}_n(x) = \sum_{i=1}^n w_i I(X_i \leq x). \tag{9}$$

It was shown in Yuan *et al.* (2012) [27] that inference using the weighted empirical distribution function is more efficient than that using the empirical distribution function without the weights. Also, since there are no closed form expressions for the  $w_i$ 's, they showed the following formula, with  $\Omega = E[g(X)g'(X)]$  and assume non-singularity,

$$w_i \stackrel{a.s.}{\approx} \frac{1}{n} \left( 1 - g(x_i)\Omega^{-1} \frac{1}{n} \sum_{k=1}^n g(x_k) \right) + O(n^{-1} \log \log(n)). \tag{10}$$

Thus, computationally, the  $w_i$ 's are well approximated by the first term in the right hand side above.

Now for our problem, denote  $x = (x_1, \dots, x_d)'$ , and  $x_{-j}$  be  $x$  with  $x_j$  removed. Recall the notations  $\mu, \tilde{\mu}_l, \tilde{\mu}_{l,j}^0$  and  $\tilde{\mu}_{l,j}^1$  given in Section 2.2. Let

$$\bar{X}_{lj} = n_{lj}^{-1} \sum_{i=1}^{n_{lj}} x_{ij}, (j = 1, \dots, d; l = 1, \dots, k),$$

$\tilde{\mu}_{lj} = \mu_{lj} + n_{lj}^2 / (n_{kj} n_l) D_l^{-1} (\bar{S}_l - \bar{\mu}_l)$ ,  $D_l = \sum_{j=1}^d n_{lj}^3 / (n_{kj} n_l^2)$ ,  $\tilde{\sigma}_j^2 = n_{lj} / n_{kj} - n_{lj}^4 / (n_{kj} n_l)^2 D_l^{-1}$ . Let

$$g_{lj}^0(x) = x - \tilde{\mu}_{lj}^0, (j = 1, \dots, d; l = 1, \dots, k).$$

Since

$$E[g_{lj}^0(x_j | x_{-j})] = \mathbf{0} \text{ then } E[g_{lj}^0(x_j)] = E[E[g_{lj}^0(x_j | x_{-j})]] = \mathbf{0}.$$

Let

$$\Omega_{lj}^0 = E[g_{lj}^0(X_j)]^2 \hat{\Omega}_{lj}^0 = \frac{1}{n_{lj}^0} \sum_{i=1}^{n_{lj}^0} (x_{ij} - \tilde{\mu}_{lj}^0)^2.$$

Let  $w_{lj,i}^0$  be the weights obtained via  $g_{lj}(x)$  given above, based on the responses  $\{x_{ij}; i = 1, \dots, n_{lj}^0\}$ , and define the weighted empirical distribution function for the  $j$ -th disease at stage  $l$ , for the standard treatment group, by

$$\tilde{F}_{lj}(x) = \sum_{i=1}^{n_{lj}^0} w_{lj,i}^0 I(x_{ij} \leq x). \tag{11}$$

$(j = 1, \dots, d; l = 1, \dots, k).$

Similarly, let

$$g_{lj}^1(y) = y - \tilde{\mu}_{lj}^1, (j = 1, \dots, d; l = 1, \dots, k).$$

Since

$$E[g_{lj}^1(y_j | y_{-j})] = \mathbf{0} \text{ then } E[g_{lj}^1(y_j)] = E[E[g_{lj}^1(y_j | y_{-j})]] = \mathbf{0}.$$

Let

$$\Omega_{lj}^1 = E[g_{lj}^1(Y_j)]^2 \hat{\Omega}_{lj}^1 = \frac{1}{n_{lj}^1} \sum_{i=1}^{n_{lj}^1} (y_{ij} - \tilde{\mu}_{lj}^1)^2.$$

Let  $w_{ij,i}^1$  be the weights obtained via  $g_{lj}^1(y)$  given above, based on the responses  $\{y_{ij}: i = 1, \dots, n_{lj}^1\}$ , and define the weighted empirical distribution function for the  $j$ -th disease at stage  $l$ , for the new treatment group, by

$$\tilde{G}_{lj}(y) = \sum_{i=1}^{n_{lj}^1} w_{ij,i}^1 I(y_{ij} \leq y). (j = 1, \dots, d; l = 1, \dots, k). \tag{12}$$

Then we define the nonparametric ROC curve for basket trial, for the  $j$ -th disease at stage  $l$ , based on the weighted empirical distribution functions, as

$$\tilde{ROC}_{lj}(u) = 1 - \tilde{G}_{lj}(\tilde{F}_{lj}^{-1}(1 - u)), 0 < u < 1. (j = 1, \dots, d; l = 1, \dots, k). \tag{13}$$

In contrast, the nonparametric ROC curve for the classical trial, for the  $j$ -th disease at stage  $l$ , based on the weighted empirical distribution functions

$$ROC_{lj}(u) = 1 - G_{lj}(F_{lj}^{-1}(1 - u)), 0 < u < 1. (j = 1, \dots, d; l = 1, \dots, k),$$

with  $G_{lj}$  and  $F_{lj}$  being the empirical distribution functions based on the data  $\{y_{ji}: i = 1, \dots, n_{lj}^1\}$  and  $\{x_{ji}: i = 1, \dots, n_{lj}^0\}$  respectively. Since  $\tilde{ROC}_{lj}(\cdot)$  is constructed with auxiliary information from data across all the diseases, it is expected to have better performance than its counterpart  $ROC_{lj}(\cdot)$  without such information.

### 3. SIMULATION STUDY

In this section, we conduct simulation studies to compare the ROC curves for basket sequential clinical trial vs the classical sequential trial, for both the parametric and nonparametric cases.

#### 3.1. Simulation Study for Parametric ROC

We first give results for the parametric case. Denote  $\mathbf{n}_l^0 = (n_{l1}^0, \dots, n_{ld}^0)$  and similarly for  $\mathbf{n}_l^1 = (n_{l1}^1, \dots, n_{ld}^1)$ , ( $l = 1, \dots, k$ ). We simulated the data for  $d = 5$  diseases, for  $k = 2$  stages. The mean values and sample sizes are given in Table 1 below.

Here  $\mu_{1,a}$  represents the mean response vector of treatment arm with small gap comparing with the mean response of control group  $\mu_0$ . Similarly,  $\mu_{1,b}$  is the mean vector with moderate gap, and  $\mu_{1,c}$  denotes mean vector with largedifference between treatment group and control group. In the fifth row,  $\mathbf{n}_1^1$  represents the sample size vector of treatment arm in stage one and  $\mathbf{n}_1^0$  be the sample size vector of control arm in stage one in the sixth row. Analogously,  $\mathbf{n}_2^1$  and

$\mathbf{n}_2^0$  denote the sample size vectors of treatment group and control group in stage two respectively.

**Table 1: Simulation Setup for 5 Disease**

Disease	1	2	3	4	5
$\mu_{1,a}$	1.3	1.0	1.6	1.2	1.4
$\mu_{1,b}$	1.8	1.5	2.1	1.6	1.9
$\mu_{1,c}$	2.5	1.9	2.9	2.5	2.7
$\mu_0$	1.2	0.9	1.5	1.0	1.3
$\mathbf{n}_1^1$	10	12	8	13	9
$\mathbf{n}_1^0$	45	60	40	55	50
$\mathbf{n}_2^1$	20	24	16	26	18
$\mathbf{n}_2^0$	100	110	95	105	100

The ROC's for the data in Table 1 are plotted in Figures 1-3. Real line is the ROC for basket trail, the upper dashed line is the ROC for independent trial. The diagonal line is drawn for reference only.

Figure 1 shows the ROC's of basket and classical trails for the simulated five diseases, with large mean differences (gap) for case and control, the AUC is computed in the bracket below, with number on the left be the AUC for the basket trail, and on the right for that of the classical trial. We see that the ROC of the former is apparently higher overall, and the AUC is larger.

Figure 2 shows the ROC's of basket and classical trails for the simulated five disease, with small mean difference. In this case the ROC's for the two types of trails are close and both not significant.

Figure 3 shows the ROC's of the two types pf trails for the simulated five disease, with moderate mean difference. In this case we see from the ROC's that the advantage of the basket trial is apparent.

Overall, according to the three figures for five diseases above, we can witness that the basket design, with side information from all the disease data, performs better than the traditional design, this can be indicated by larger AUC. Meanwhile, the plotwith large gap shows the best result among the three, which means as the gap between treatment arm and control group increasing, the result becomes more apparent. Generally speaking, ROC curves behave better in stage one than stage two in all three plots. More specifically, based on Figure 1, the largest distance between AUC is about 0.166 and the smallest one is about 0.011. From Figure 2, the biggest gap between AUC is 0.032. On the contrary, the closest distance is about 0.002. According to Figure 3, the widest difference is about 0.101 and the most narrow one is about 0.002.

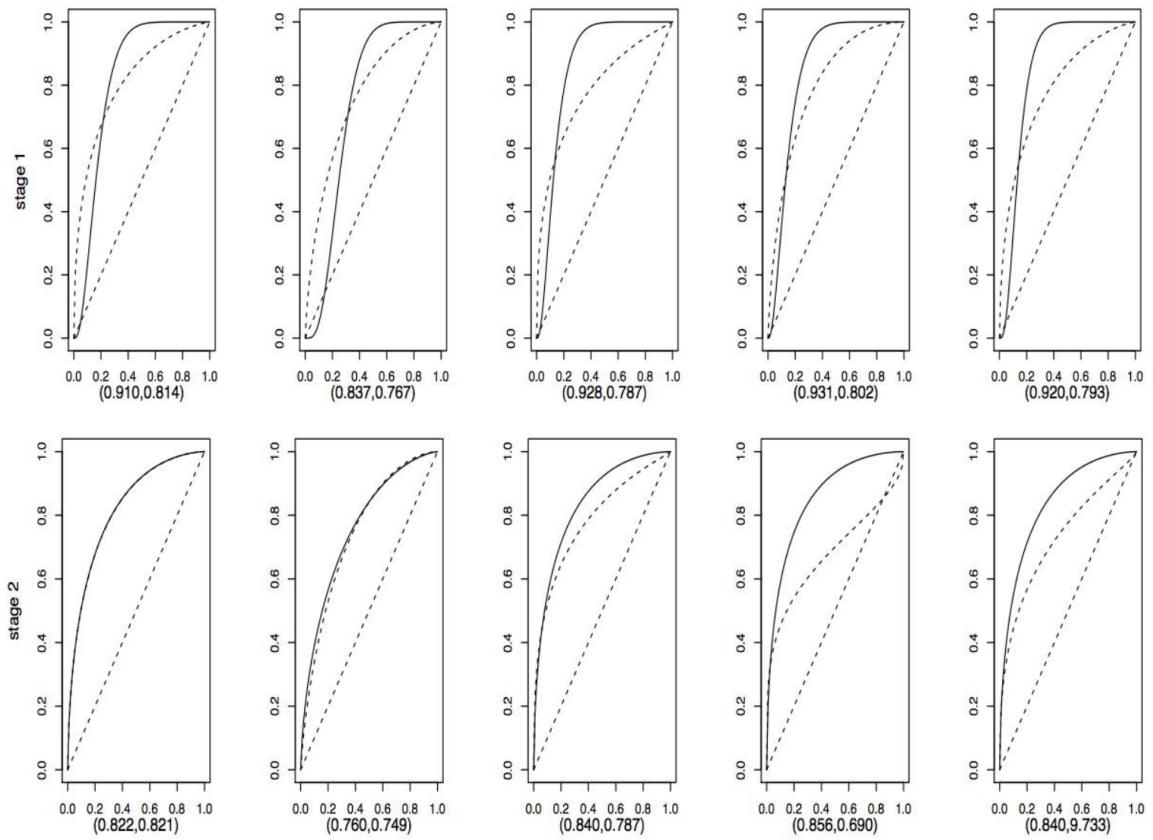


Figure 1: Parametric ROC Curve for 5 simulated diseases (large gap). Upper row: stage I; lower row: stage II.

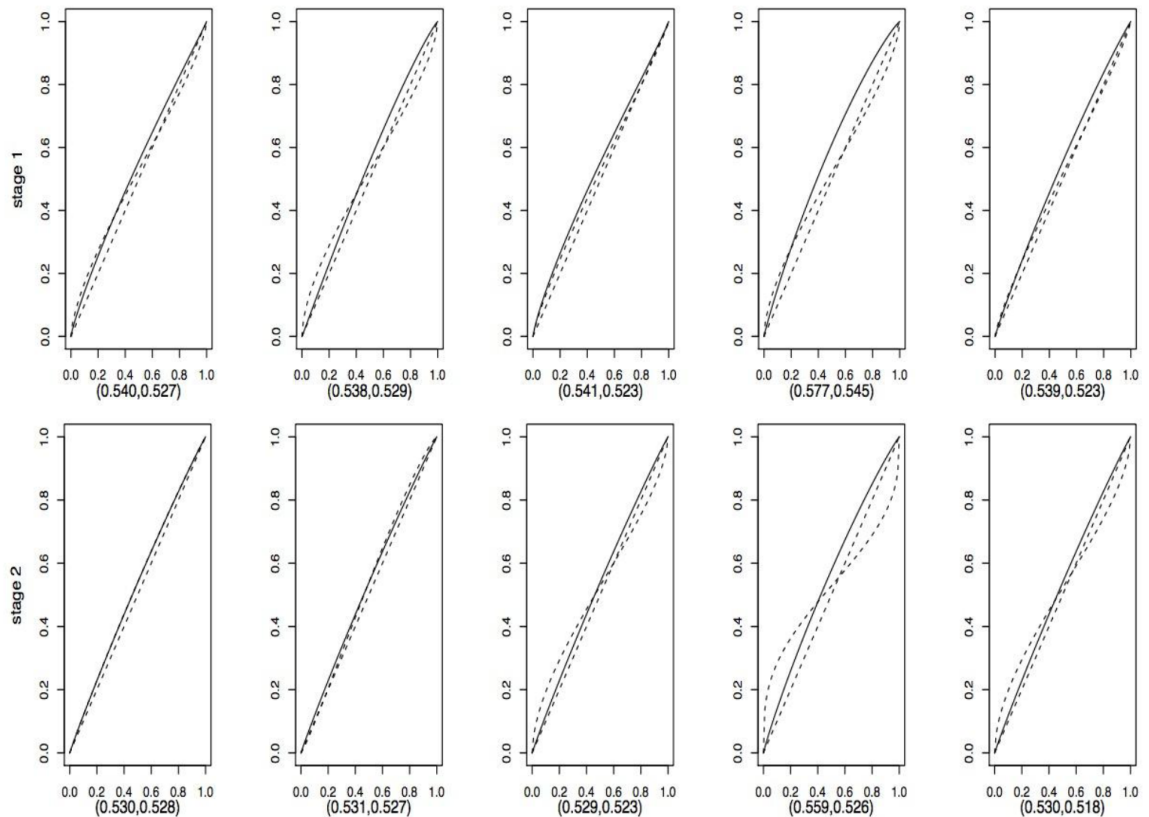


Figure 2: Parametric ROC Curve for 5 simulated diseases (small gap). Upper row: stage I; lower row: stage II.

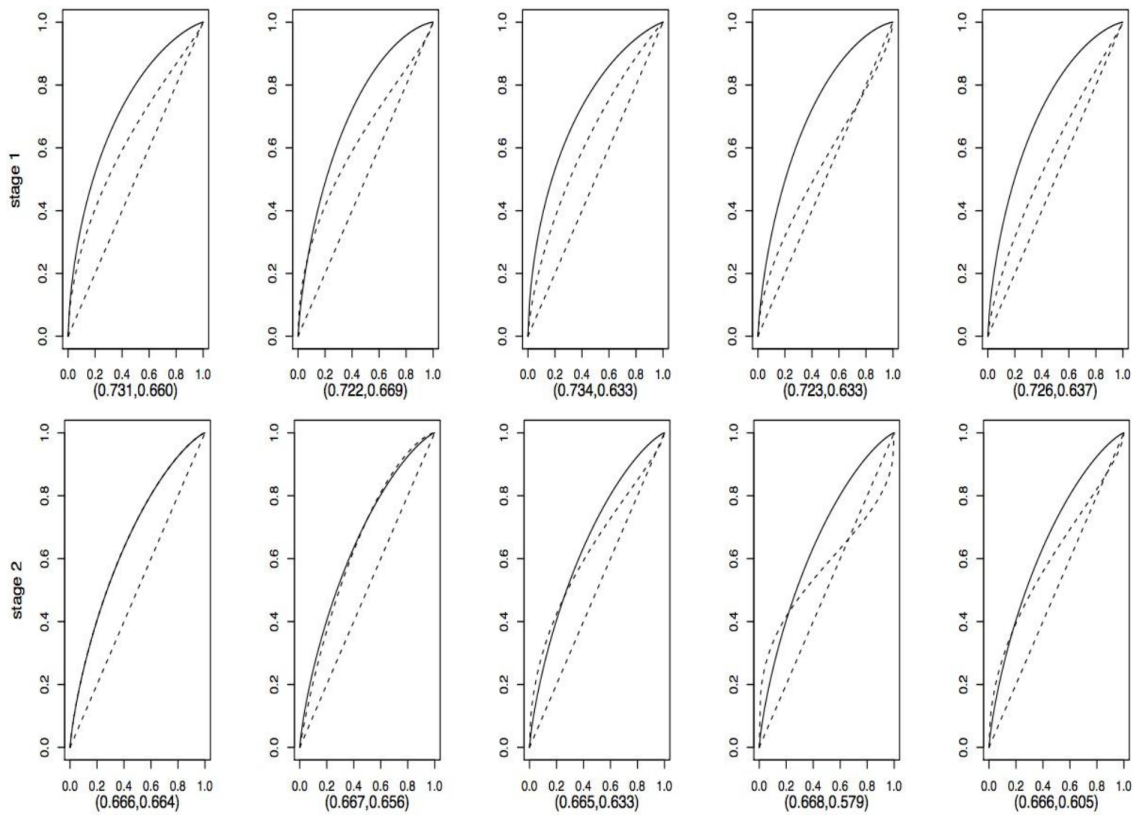


Figure 3: Parametric ROC Curve for 5 simulated diseases (moderate gap). Upper row: stage I; lower row: stage II.

Below we perform a simulation with nine diseases, the setup is shown in Table 2.

Table 2: Simulation Data for 9 Disease

Disease	1	2	3	4	5	6	7	8	9
$\mu_1$	2.5	1.9	2.9	2.5	2.7	3.1	2.8	2.4	2.6
$\mu_0$	1.2	0.9	1.5	1.0	1.3	1.7	1.4	1.1	1.2
$n_1^1$	10	12	8	13	9	10	11	7	10
$n_1^0$	45	60	40	55	50	49	50	51	52
$n_2^1$	20	24	16	26	18	20	22	14	20
$n_2^0$	100	110	95	105	100	98	100	102	104

The corresponding ROC's are show in Figures4-5. We see that the advantage of basket trial over the classical one is apparent. The conclusions are similar to the case of five diseases.

### 3.2. Simulation Study for Nonparametric ROC.

Now we use the setup in Table 1 to examine the performance of the nonparametric ROC's. In Figure 6, for a fixed disease, we compare the empirical ROC (dashed line step function) with the parametric normal ROC (dotted line) and the true ROC (solid line) obtained from the true underlying distributions. The sample sizes for case and control are given below the figures. We see that for small sample size  $(n_1, n_0) = (10, 45)$ , the advantage of the empirical ROC is not clear. As the sample size increases, the empirical one becomes

more and more close to the real one, while the ROC from normal model yields biased estimate of the real ROC.

In Figure 7, we compare the empirical ROC's (dashed line) for the basket trail with side information incorporated, given in (10)-(12), to the empirical ROC (solid line) from the classical trial. Again, the ROCs from the basket trail out performs those of the classical trial.

### 3.3. A Hypothetical Basket Clinical Trial

A basket clinical trial has been conducted by Hyman, et al. (2015) [28] to evaluate the efficacy of a therapeutic oral treatment, Vemurafenib. The drug is an inhibitor of BRAF V600E enzyme which can improve the survival rate in melanoma patients whose cancer has a V600E BRAF mutation. To investigate the treatment effect on nonmelanoma cancer patients, the trial enrolled 122 patients with nonmelanoma cancers. The study included patients with non-small-cell lung cancer (NSCLC), ovarian cancer, colorectal cancer, cholangiocarcinoma, breast cancer, and multiple myeloma, and cancer due to any other BRAF V600 mutation. Among these groups, the three commonly occurred disease groups are colorectal cancer (CC) with  $n_1 = 37$  patients, NSCLC with  $n_2 = 20$  patients, Erdheim-Chester disease (ECD) or Langerhans cell histiocytosis (LCH) with  $n_3 = 18$  patients.

The original trial used the survival rates as patients' final outcomes. The outcomes such as the percentage

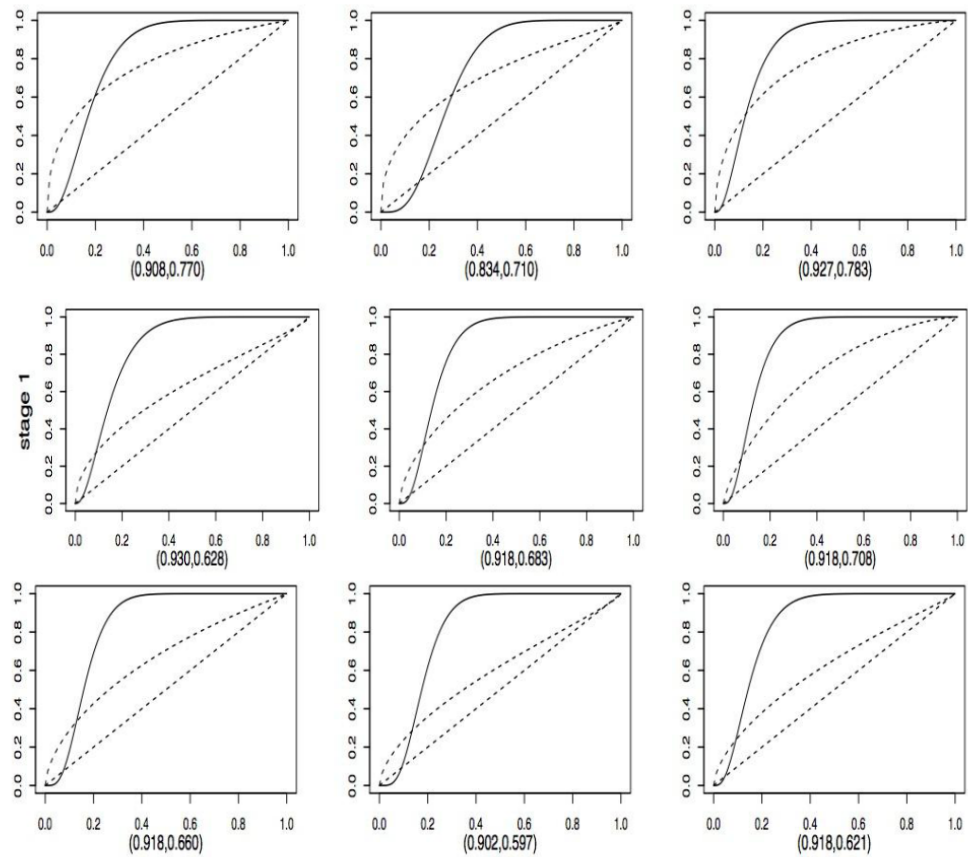


Figure 4: Parametric ROC Curve, Stage I for 9 simulated diseases.

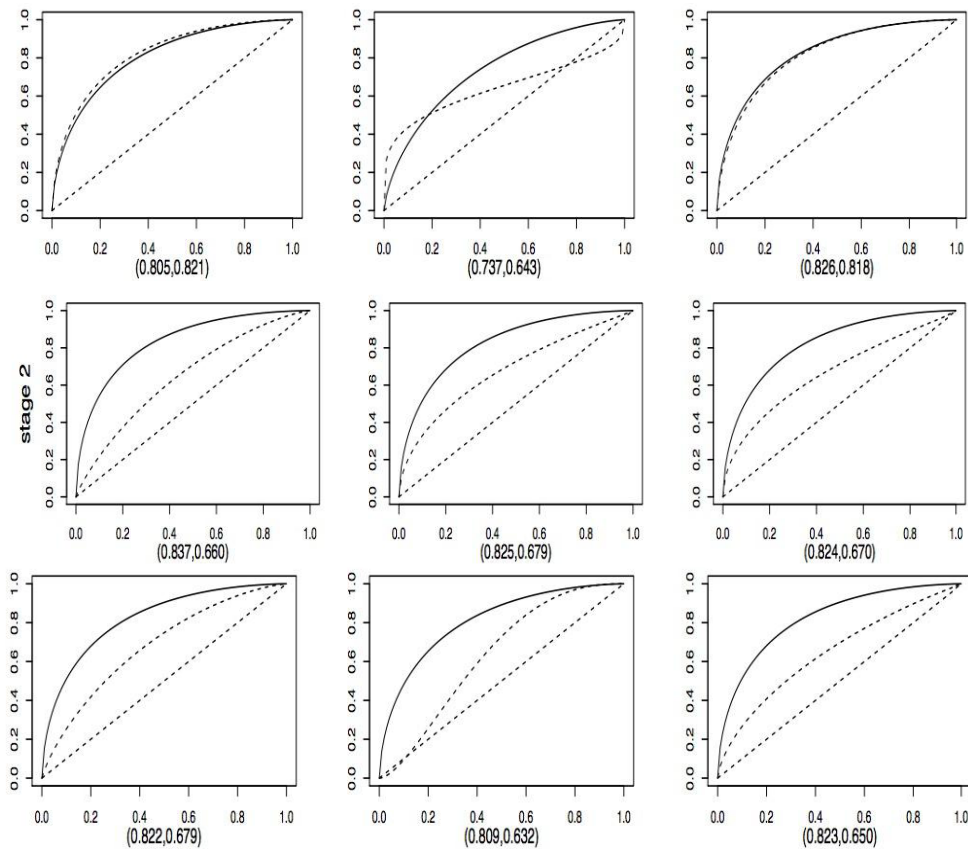
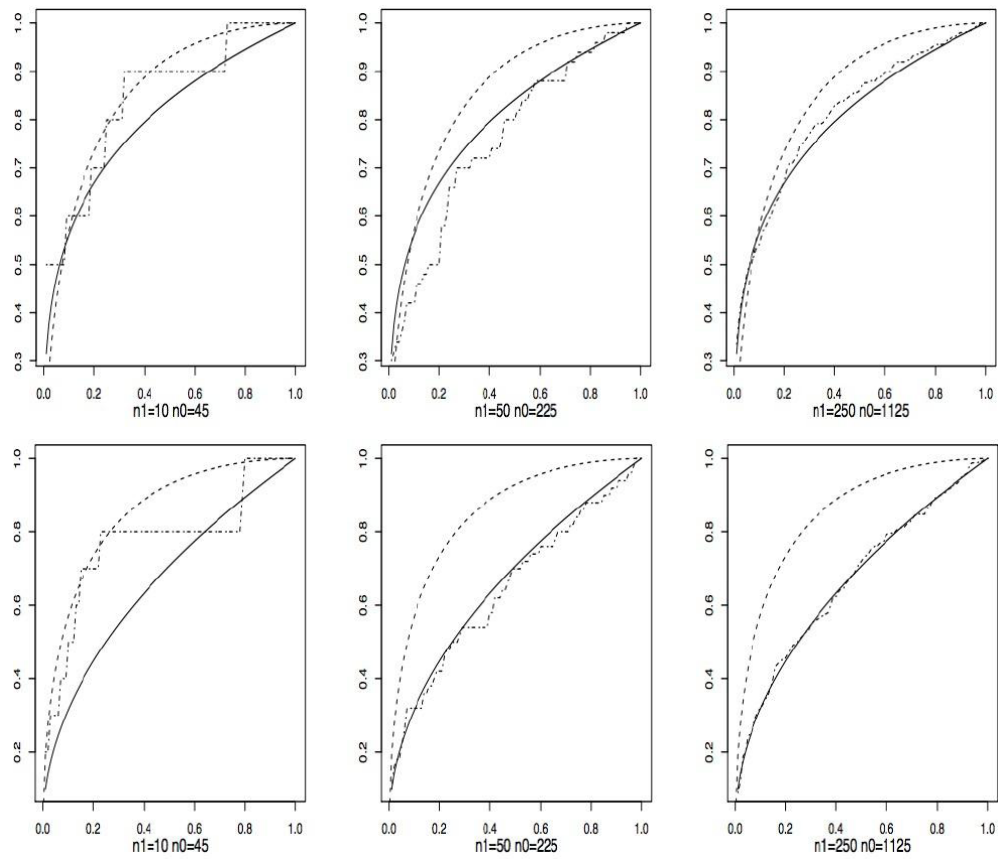
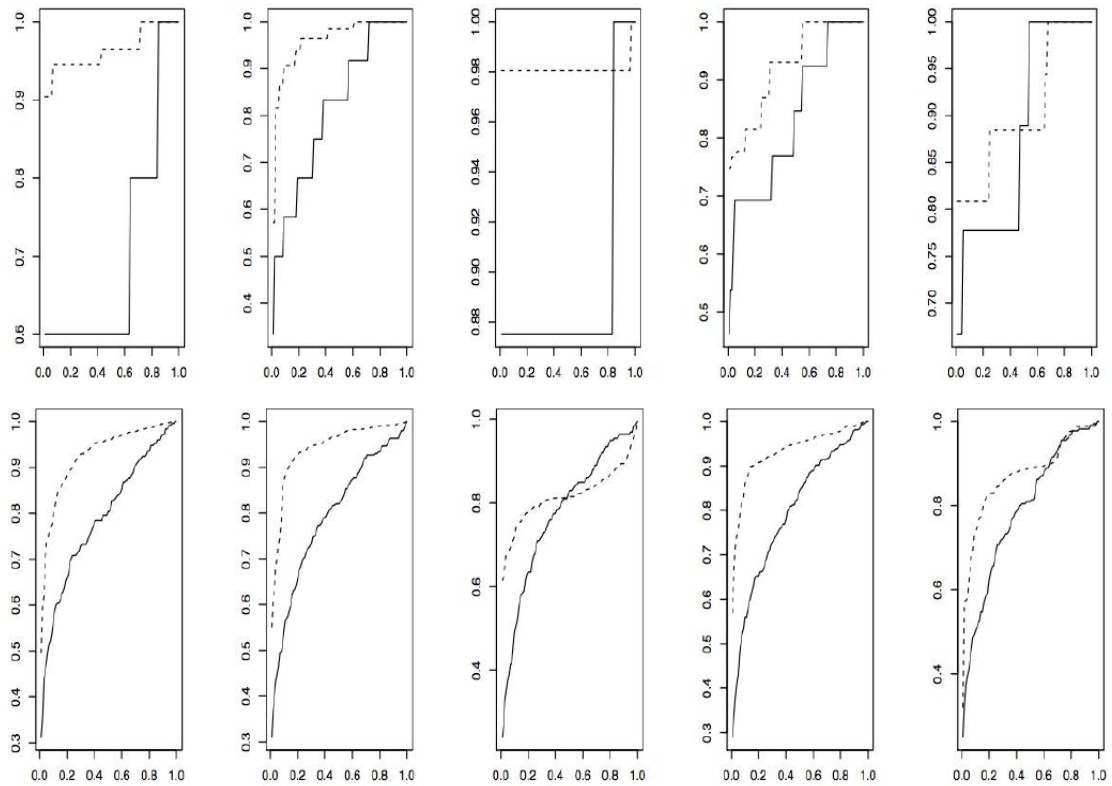


Figure 5: Parametric ROC Curve, Stage II for 9 simulated diseases.





**Figure 6:** Nonparametric ROC Curve, Stage I for 5 simulated diseases.



**Figure 7:** Nonparametric ROC Curve, Stage II for 5 simulated diseases.

of tumor shrinkage may capture more information so that the trial may be better powered. Since the normal approximation is reasonable for percentages, we use the percentage of tumor shrinkage as patients' outcomes in this hypothetical example. According to a retrospective review in Dadu, *et al.* (2014) [29], most patients treated by Vemurafenib had tumor shrinkage ranging from 3% to 58%. We consider the three most common groups with the pre-specified samples sizes. Here  $d = 3$ . We assume a two-stage group sequential design, and equal allocation between treatment and control for each disease. In the first stage, tumor shrinkage is observed from  $n_{11}^1 = 10$  CC patients treated by Vemurafenib, and  $n_{11}^0 = 9$  control CC patients. Similarly, the numbers of patients observed at the first stage from other two diseases are  $n_{12}^1 = 5$  and  $n_{12}^0 = 5$  for NSCLC,  $n_{13}^1 = 5$  and  $n_{13}^0 = 5$  for ECD or LCH. The numbers of patients at the second stage for these three diseases are  $n_{21}^1 = 9$  and  $n_{21}^0 = 8$  for CC,  $n_{22}^1 = 5$  and  $n_{22}^0 = 5$  for NSCLC, and  $n_{23}^1 = 4$  and  $n_{23}^0 = 4$  for ECD or LCH. We are interested in testing the null hypothesis

$$H_0: \mu \leq \mu_0, \text{ vs } H_1: \mu > \mu_1 = \mu_0 + \delta,$$

where the mean tumor shrinkage  $\mu_0 = (0.10, 0.20, 0.58)'$  is assumed from aforementioned range on patients treated by Vemurafenib. Here we let  $\delta = 0$ . The multivariate normal distribution of the tumor size shrinkage is assumed to be

$$X \sim N(\beta, \Omega).$$

Take  $\theta = 0$ , then by the formula in the Appendix,  $\beta = \mu_0$ . The diagonal elements in the covariance matrix,  $\Omega$ , calculated from the variance expression for percentages, are (0.3, 0.4, 0.5) for three diseases. The correlation is assumed to be 0.1 for three diseases within the same patient. The hypothetical example dataset was simulated from this multivariate normal distribution.

With the normal distribution assumption, we apply the parametric ROC method to the simulated dataset. To borrow information from all the disease types, the values for  $\tilde{S}_l^0 = (0.121, 0.247)$  and  $\tilde{\mu}_l^0 = (0.253, 0.242)$ , are obtained from the simulated dataset. The conditional statistic has the distribution:

$$S_{l,j}^0 | \tilde{S}_l^0 \sim N(\tilde{\mu}_j^0, (\tilde{\sigma}_{lj}^0)^2),$$

where  $\tilde{\mu}_j^0 = (-0.221, 0.041, 0.382)$  ( $l = 1$ ),  $(-1.080, -0.261, 0.285)$  ( $l = 2$ ),  $D_{l0} = (0.408, 0.363)$ , and  $(\tilde{\sigma}_{lj}^0)^2 = (1.009, 0.972, 1.206)$  ( $l = 1$ ),  $(0.995, 0.987, 0.980)$  ( $l = 2$ ). The level  $\alpha = 0.05$  cut-off point is obtained accordingly. Let  $n_l^1 = (20, 18)$ ,  $\tilde{S}_l^1 = (0.332, 0.014)$  and  $\tilde{\mu}_l^1 = (1.60, 1.589)$ ,

$$S_{l,j}^1 | \tilde{S}_l^1 \sim N(\tilde{\mu}_j^1, (\tilde{\sigma}_{lj}^1)^2),$$

where  $\tilde{\mu}_j^1 = (1.287, 1.676, 2.070)$  ( $l = 1$ ),  $(1.224, 1.658, 2.066)$  ( $l = 2$ ),  $D_{l1} = (0.418, 0.377)$ , and  $(\tilde{\sigma}_{lj}^1)^2 = (0.982, 0.974, 1.209)$  ( $l = 1$ ),  $(0.906, 0.971, 0.981)$  ( $l = 2$ ). The parametric ROC curves under normal model for the three diseases at the first and second interim looks are given by

$$ROC_{l,j}(u) = \Phi\left(\frac{\tilde{\mu}_{lj}^1 - \tilde{\mu}_{lj}^0}{\tilde{\sigma}_{lj}^1} + \tilde{\sigma}_{lj}^0 \tilde{\sigma}_{lj}^1 \Phi^{-1}(u)\right), 0 < u < 1, (j = 1, 2, 3; l = 1, 2).$$

We also illustrate the application of our nonparametric method. The ROC curves are constructed using the empirical distribution functions. Let

$$g_{lj}^0(x) = x - \tilde{\mu}_{lj}^0, (j = 1, \dots, d; l = 1, \dots, k),$$

where  $\tilde{\mu}_{lj}^0 = (-0.317, -0.006, 0.322)$  is given in the previous paragraph. The weights  $w_{lj,i}^0$  are obtained via  $g_{lj}(x)$ , based on the simulated dataset, and the weighted empirical distribution function for the  $j$ -th disease at stage  $l$ , for the control group, by

$$\tilde{F}_{lj}(x) = \sum_{i=1}^{n_{lj}^0} w_{lj,i}^0 I(x_{j,i} \leq x). (j = 1, \dots, d; l = 1, \dots, k),$$

where 870, 0.1731737, 0.2762839, 0.2076634 ( $j=2$ ), (0.003823371 0.041840711 0.170644613  $w_{lj,i}^0 = (0.107, 0.106, 0.120, 0.125, 0.134, 0.055, 0.143, 0.127, 0.093)$  ( $j = 1$ ), (0.104, 0.239, 0.173, 0.276, 0.208) ( $j = 2$ ), (0.004, 0.042, 0.171, 0.067, 0.717) ( $j = 3$ ) according to Equation (10). For the Vemurafenib group,

$$g_{lj}^1(y) = y - \tilde{\mu}_{lj}^1, (j = 1, \dots, d; l = 1, \dots, k),$$

where  $\tilde{\mu}_{lj}^1 = (1.225005, 1.666252, 2.057815)$ . The weights are obtained to be  $w_{lj,i}^1$ , and the weighted empirical distribution function for the  $j$ -th disease at stage  $l$ , for the Vemurafenib group, by

$$\tilde{G}_{lj}(y) = \sum_{i=1}^{n_{lj}^1} w_{lj,i}^1 I(y_{j,i} \leq y). (j = 1, \dots, d; l = 1, \dots, k),$$

where 1), (0.35553318 0.17883773 0.10849838 0.28989895 0.06723176) ( $j=2$ ),  $w_{lj,i}^1 = (0.003, 0.043, 0.051, 0.116, 0.010, 0.291, 0.071, 0.015, 0.235, 0.166)$  ( $j = 1$ ), (0.356, 0.179, 0.108, 0.290, 0.067) ( $j = 2$ ), (0.094, 0.136, 0.158, 0.357, 0.254) ( $j = 3$ ), according to Equation (10).

Based on these estimated distribution functions, the six nonparametric ROC curve for the  $j$ -th disease at stage  $l$ , based on the weighted empirical distribution functions, are given by

$$ROC_{l,j}(u) = 1 - \tilde{G}_{lj}(\tilde{F}_{lj}^{-1}(1 - u)), 0 < u < 1. (j = 1, 2, 3; l = 1, 2).$$

**CONCLUDING REMARKS**

We evaluated the performances of basket trial and the classical trial by comparing their ROCs, with both parametric normal model and nonparametric model. Simulation studies show that the former is generally better in terms of higher ROCs. We find that when the total sample sizes for case and control are small, say no bigger than (100,100), the parametric method is more preferred; when the sample size is (100,100) or bigger, the nonparametric ROC gives better fit to the true ROC, while that from the normal model can give biased result.

**ACKNOWLEDGEMENT**

This work is supported in part by the Intramural Research Program of the National Institutes of Health.

**APPENDIX**

**Derivation of (2).** Since  $P(X_j|C = c) = N(\mu_0 + c\eta_j, \sigma_j^2)$  and  $C \sim N(\theta, \tau^2)$ ,

$$\begin{aligned}
 P(X) &= \int \prod_{j=1}^d P(X_j|C = c)P(c)dc \\
 &= \frac{1}{(2\pi)^{(d+1)/2}\tau\sigma_1 \dots \sigma_d} \int \exp\left\{-\frac{1}{2} \sum_{j=1}^d \frac{(x_j - \mu_0 - \eta_j c)^2}{\sigma_j^2} - \frac{1}{2} \frac{(c - \theta)^2}{\tau^2}\right\} dc \\
 &= \frac{1}{(2\pi)^{(d+1)/2}\tau\sigma_1 \dots \sigma_d} \int \exp\left\{-\frac{1}{2} \left(\sum_{j=1}^d \frac{\eta_j^2 c^2 - 2\eta_j(x_j - \mu_0)c + (x_j - \mu_0)^2}{\sigma_j^2} + \frac{c^2 - 2c\theta + \theta^2}{\tau^2}\right)\right\} dc \\
 &\propto \int \exp\left\{-\frac{\sum_{j=1}^d (a_j \eta_j^2 c^2 - 2a_j \eta_j (x_j - \mu_0)c + a_j (x_j - \mu_0)^2) + a_0 c^2 - 2ca_0 \theta + a_0 \theta^2}{\tau^2 \sigma_1^2 \dots \sigma_d^2}\right\} dc \\
 &\propto \int \exp\left\{-\frac{c^2 - 2\frac{a_0 \theta + \sum_{j=1}^d a_j \eta_j (x_j - \mu_0)}{b} c}{2a_0 \tau^2 / b}\right\} \exp\left\{-\frac{\sum_{j=1}^d a_j (x_j - \mu_0)^2}{2a_0 \tau^2 / b}\right\} dc \\
 &\propto \int \exp\left\{-\frac{(c - e)^2}{2a_0 \tau^2 / b}\right\} \exp\left\{-\frac{\sum_{j=1}^d a_j (x_j - \mu_0)^2 - e^2}{2a_0 \tau^2 / b}\right\} dc \\
 &\propto \exp\left\{-\frac{\sum_{j=1}^d a_j (x_j - \mu_0)^2 - e^2}{2a_0 \tau^2 / b}\right\},
 \end{aligned}$$

where,  $a_0 = \sigma_1^2 \dots \sigma_d^2$ ,  $a_j = \tau^2 \prod_{i \neq j} \sigma_i^2$ ,  $b = a_0 + \sum_{j=1}^d a_j \eta_j^2$  and  $e = (a_0 \theta + \sum_{j=1}^d a_j \eta_j (x_j - \mu_0)) / b$ . Let  $\Omega^{-1} = (\omega_{ij})_{d \times d}$ , with

$$\omega_{jj} = a_j (b^2 - \eta_j^2) / (a_0 \tau^2 b), \omega_{ij} = a_i a_j \eta_i \eta_j / (a_0 \tau^2 b), (i \neq j).$$

Assume  $b^2 - \eta_j^2 > 0$  for  $j = 1, \dots, d$ , let  $s = (s_1, \dots, s_d)'$  with  $s_j = a_j \eta_j b \theta / 2$ , and  $\mathbf{1} = (1, \dots, 1)'$ , a  $d$ -vector of 1's. Then

$$\begin{aligned}
 &\exp\left\{-\frac{\sum_{j=1}^d a_j (x_j - \mu_0)^2 - e^2}{2a_0 \tau^2 / b}\right\} \\
 &\propto \exp\left\{-\frac{1}{2} ((x - \mu_0 \mathbf{1})' \Omega^{-1} (x - \mu_0 \mathbf{1}) - 2s'(x - \mu_0 \mathbf{1}))\right\} \\
 &= \exp\left\{-\frac{1}{2} ((\Omega^{-1/2} (x - \mu_0 \mathbf{1}))' (\Omega^{-1/2} (x - \mu_0 \mathbf{1}) - 2(\Omega^{1/2} s)' (\Omega^{-1/2} (x - \mu_0 \mathbf{1})))\right\} \\
 &\propto \exp\left\{-\frac{1}{2} ((\Omega^{-1/2} (x - \mu_0 \mathbf{1} - \Omega s))' (\Omega^{-1/2} (x - \mu_0 \mathbf{1} - \Omega s)))\right\} \\
 &= \exp\left\{-\frac{1}{2} ((x - \mu_0 \mathbf{1} - \Omega s)' \Omega^{-1} (x - \mu_0 \mathbf{1} - \Omega s))\right\},
 \end{aligned}$$

Thus,  $X \sim N(\beta, \Omega)$ , with  $\beta = \mu_0 \mathbf{1} + \Omega s$ .

**Derivation of (4).** Note that

$$\begin{aligned}
 \begin{pmatrix} S_{lj} \\ \bar{S}_l \end{pmatrix} &\sim N(\tilde{\mu}, \tilde{\Omega}), \tilde{\mu} = \begin{pmatrix} \mu_{lj} \\ \bar{\mu}_l \end{pmatrix}, \tilde{\Omega} \\
 &= \begin{pmatrix} n_{lj}/n_{kj} & n_{lj}^2/(n_{kj}n_l) \\ n_{lj}^2/(n_{kj}n_l) & D_l \end{pmatrix},
 \end{aligned}$$

where  $\mu_{lj}$  is the  $j$ -th entry of  $\mu_{lk} = \Omega^{-1/2}(\mu_{01}n_{l1}/\sqrt{n_{k1}} \dots, \mu_{0d}n_{ld}/\sqrt{n_{kd}})'$ ,  $\bar{\mu}_l = \sum_{j=1}^d \mu_{lj} n_l / n_l$ , and  $D_l = \sum_{j=1}^d n_{lj}^2 / (n_{kj} n_l^2)$ . Thus,  $S_{lj} | \bar{S}_l \sim N(\tilde{\mu}_j, \tilde{\sigma}_j^2)$ , with  $\tilde{\mu}_j = \mu_{lj} + n_{lj}^2 / (n_{kj} n_l) D_l^{-1} (\bar{S}_l - \bar{\mu}_l)$  and  $\tilde{\sigma}_j^2 = n_{lj} / n_{kj} - n_{lj}^2 / (n_{kj} n_l) D_l^{-1}$ .

**REFERENCES**

- [1] Willyard C. Basket studies?will hold intricate data for cancer drug approvals. Nature Medicine 2013; (6): 655. <https://doi.org/10.1038/nm0613-655>
- [2] Redig AJ, Janne PA. Basket trials and the evolution of clinical trial design in an era of genomic medicine. Journal of Clinical Oncology 2015; (20): 975-977. <https://doi.org/10.1200/JCO.2014.59.8433>
- [3] Beckman RA, Antonijevic Z, Kalamegham R, Chen C. Design for a phase 3 basket trial in multiple tumor types based on a putative biomarker. Clinical Cancer Research 2015; In press.
- [4] Chen C, Li N, Yuan S, Antonijevic Z, Kalamegham R, Beckman RA. Statistical design and consideration of a phase 3 basket trial for simultaneous investigation of multiple tumor types in one study 2016.
- [5] Barker AD, Sigman CC, Kelloff GJ, Hylton NM, Berry DA, Esserman LJ. I-SPY 2: an adaptive breast cancer trial design in the setting of neoadjuvant chemotherapy. Clinical Pharmacology & Therapeutics 2009; 97-100. <https://doi.org/10.1038/clpt.2009.68>

- [6] Kopetz S. Right Drug for the Right Patient: Hurdles and the path forward in colorectal cancer. ASCO educational book 2013. <http://meetinglibrary.asco.org/content/19-132>.
- [7] Meador CB, Micheel CM, Levy MA, Lovly CM, Horn L, Warner JL, *et al.* Beyond histology: translating tumor genotypes into clinically effective targeted therapies. *Clinical Cancer Research* 2014; 20: 2264-2275.
- [8] Lacombe D, Burocka S, Bogaerts J, Schoeffskib P, Gofinopoulousa V, Stuppa R. The dream and reality of histology agnostic cancer clinical trials. *Molecular Oncology* 2014; 8: 1057-1063.
- [9] Sleijfer S, Bogaerts J, Siu LL. Designing transformative clinical trials in the cancer genome era. *Journal of Clinical Oncology* 2013; 31: 1834-1841.
- [10] Demetri G, Becker R, Woodcock J, Doroshow J, Nisen P, Sommer J. Alternative trial designs based on tumor genetics/pathway characteristics instead of histology. Issue Brief: Conference on Clinical Cancer Research; <http://www.focr.org/conference-clinical-cancer-research> 2011.
- [11] Metz CE. Basic principles of ROC analysis. In *Seminars in Nuclear Medicine* 1978; (4): 283-298. WB Saunders. [https://doi.org/10.1016/s0001-2998\(78\)80014-2](https://doi.org/10.1016/s0001-2998(78)80014-2)
- [12] Zhou XH, Obuchowski AA, McClish DK. *Statistical Methods in Diagnostic Medicine*. 2011; Vol. 712: Wiley. com. <https://doi.org/10.1002/9780470906514>
- [13] Tang LL, Zhou XH. Semiparametric separation curve approach for comparing correlated ROC data from multiple markers, *Journal of Computational and Graphical Statistics* 2012; 662-676. <https://doi.org/10.1080/10618600.2012.663303>
- [14] Tang LL, Yuan A, Collins J, Che X, Chan L. Unified least squares methods for the evaluation of diagnostic tests with the gold standard 2016.
- [15] Pocock SJ. Group sequential methods in the design and analysis of clinical trials. *Biometrika* 1977; 191-199. <https://doi.org/10.1093/biomet/64.2.191>
- [16] O'Brien PC, Fleming TR. A multiple testing procedure for clinical trials. *Biometrics* 1979; 549-556. <https://doi.org/10.2307/2530245>
- [17] Berry DA. Interim analysis in clinical trials: Classical vs. Bayesian approaches. *Statistics in Medicine* 1985; 521-26. <https://doi.org/10.1002/sim.4780040412>
- [18] Wang SK, Tsaitis AA. Approximately optimal one-parameter boundaries for group sequential trials. *Biometrics* 1987; 193-200. <https://doi.org/10.2307/2531959>
- [19] Moss AJ, Hall WJ, Cannon DS, *et al.* Improved survival with an implanted defibrillator in patients with coronary disease at high risk for ventricular arrhythmia. *The New England Journal of Medicine* 1996; 1933-1940. <https://doi.org/10.1056/NEJM199612263352601>
- [20] Bellissant E, Duhamel JF, Guillot M, *et al.* The triangular test to assess the efficacy of metoclopramide in gastroesophageal reflux. *Clinical Pharmacology and Therapeutics* 1997; 377-384. [https://doi.org/10.1016/S0009-9236\(97\)90170-3](https://doi.org/10.1016/S0009-9236(97)90170-3)
- [21] Tan M, Xiong X, Kutner MH. Clinical trial designs based on sequential conditional probability ratio tests and reverse stochastic curtailing. *Biometrics* 1998; 682-695. <https://doi.org/10.2307/3109774>
- [22] Jennison C, Turnbull BW. *Group sequential methods with applications to clinical trials*, CRC Press Inc. (Boca Raton, FL) 2000.
- [23] Hung J, Wang S-J, O'Neill R. Statistical considerations for testing multiple endpoints in group sequential or adaptive clinical trials. *Journal of Biopharmaceutical Statistics* 2007; 1201-1210. <https://doi.org/10.1080/10543400701645405>
- [24] Huang P, Tan MT. Multistage nonparametric global statistical test: a solution to the Behrens-Fisher problem for multidimensional mixed outcomes. *Statistics and Its Interface* 2016; 9(3): 343-354.
- [25] Qin J, Lawless JL. Empirical likelihood and general estimating equations. *Annals of Statistics* 1994; 22: 300-325.
- [26] Owen AB. Empirical likelihood confidence regions. *Annals of Statistics* 1990; 18: 90-120.
- [27] Yuan A, He W, Wang B, Qin G. U-statistic with side information. *Journal of Multivariate Analysis* 2012; 20-38.
- [28] Hyman DM, *et al.* Vemurafenib in multiple nonmelanoma cancers with BRAF V600 mutations. *New England Journal of Medicine* 2015; 373: 726-736.
- [29] Dadu R, Shah K, Busaidy NL, Waguespack SG, Habra MA, Ying AK, Hu MI, Bassett R, Jimenez C, Sherman SI, Cabanillas ME. Efficacy and tolerability of vemurafenib in patients with BRAFV600E-positive papillary thyroid cancer: MD Anderson Cancer Center off label experience. *The Journal of Clinical Endocrinology & Metabolism* 2014; 100(1): E77-E81.