Robust Test for Multiple Endpoints in Group Sequential Design

Ao Yuan[1,2,*], John Collins[1,3], Leighton Chan[1] and Ming T. Tan[2,*]

[1]*NIH Clinical Center, Rehabilitation Medicine Department, Bethesda, MD 20892, USA*

[2]*Department of Biostatistics, Bioinformatics and Biomathematics, Georgetown University, Washington DC 20057, USA*

[3]*Department of Rehabilitation Science, George Mason University, Fairfax, VA 22030, USA*

Abstract: Group sequential design is widely used in today's phase II/III clinical trials where testing multiple endpoints is quite often performed. In such tests, a basic requirement is to control the family-wise error rate at a given nominal level. The design is determined by a set of testing statistic and stopping boundaries (rules). Existing methods compute the stopping boundaries use Normal approximations, which work well when the true underlying data distribution is approximately Normal, but with small sample sizes the Normal approximation may not be valid. In an attempt to overcome these difficulties, we propose a robust method to compute the stopping boundaries in which it is assumed only that the data distributions are symmetric about their means. The null and alternative distributions are then constructed via the empirical distribution as well as the stopping boundaries for the given nominal level. Powers for the test statistics are obtained by bootstrap simulation, which is always valid for any sample size, and correlations between test statistics are automatically taken care of. Simulation examples are given to illustrate the proposed method.

Keywords: Family-wise error rate, Group sequential design, Multiple endpoints, Robust test, Primary endpoint, Secondary endpoint.

1. INTRODUCTION

Conducting a clinical trial is an expensive and time-consuming process. Consequently, there is value in stopping a phase II/III early if there is unambiguous evidence for or against the intervention's efficacy or if unexpected conditions require modification to the design. In such cases a group sequential (GS) design may be used. The group sequential (or multi-stage) design was pioneered [22, 25] and further studied by many (*e.g*. [1, 2, 11, 12, 21, 30, 32]). After each stage of a GS design, interim analyses are conducted and compared against a predetermined stopping rule. If results are more extreme than the stopping rule for that stage, then the trial is terminated. It is common for GS designs to have multiple outcomes of interest, and so to incorporate multiple endpoints corresponding to the outcomes into their stopping rule. The stopping rule is therefore constructed as a stage and endpoint-dependent boundary on the test statistics in order to limit the family-wise error rate (FWER) to a given level.

Existing methods for constructing stopping rules rely on assumptions that are not robust for small sample sizes. In an attempt to overcome this difficulty, we propose a robust method to compute the stopping boundaries in the case of one-armed studies. The extension to two or more arms and multiple endpoints is briefly discussed in the last section. The proposed method can be viewed as an exact test for continuous data, in which the critical boundaries are determined by the empirical distribution of the observed data, instead of asymptotic distribution.

In Section 2 we describe the existing methods for constructing stopping rules for GS designs. Section 3 provides the set-up and notation for the proposed method, which is introduced in section 4 for the treatment-only and treatment-placebo cases. Illustrative examples using the method are provided in section 5 and a brief discussion is given in the final section.

2. EXISTING METHODS

Many existing methods (*e.g*. [3, 5, 11, 16, 23, 26, 36, 37]) compute stopping boundaries using Normal distributions on the multiple test statistics. [17] Proposed a method to test hypothesis about the secondary boundary. All these methods are easy to implement and work well when the true underlying distribution is approximately Normal.

In practice, there are two problems with the Normal approximation. The sample sizes in phase II/III clinical trials are sometimes small, under which the Normal approximations for the test statistics may not be valid even when joint Normality holds for the true distributions. Even if the Normal approximation is valid, the family-wise error rate depends on unknown correlations among

*Address correspondence to: Ao Yuan and Ming T. Tan;
Tel: 202-687-6766;
E-mail: ay312@georgertown.edu; mtt34@georgetown.edu

these test statistics. In the latter case a correlation structure must be assumed. [29] investigated the behavior of stopping boundaries under normal distribution of the data and showed that for some boundaries the FWER are not controlled at the given nominal level. In fact, for many cases the actual significance level is smaller than the nominal level and the nominal level is achieved only if the primary and secondary points are in perfect positive correlation. The resulting boundaries used by methods assuming joint Normality are therefore generally conservative. Moreover, under joint Normality there are no closed form solutions for the stopping boundaries.

Naively, the drawbacks of assuming joint Normality for small samples suggest that some sort of exact test may be a better approach. The exact test for contingency table data proposed by [6] is popular for small sample sizes. It has been extensively studied and extended to various types of discrete data (*e.g.* [4, 8, 9, 19, 28, 35]). [19] provide a rather complete collection of results and recent developments in this area that includes an extension of discrete results to continuous data. For most continuous data problems, the data are either converted into ranks or discretized to use these methods, which leads to reduced efficiency as not all information is being used. In only a few methods are the original continuous data used in such tests, such as the Kolmogorov-Smirnov test. These methods obtain the p-value is obtained by Monte Carlo sampling from the empirical distribution. Unfortunately, these methods do not apply to the case of location parameters under the null and alternative hypothesis.

Another way to implement null and alternative parameters for the stopping boundary into the empirical distribution is via empirical likelihood (*e.g.* [24, 27]). This method does not require any assumptions about the data distribution and is more robust, but the effect of the parameter implemented this way via empirical weights is not clear. Nor does this method lead to closed form construction of the distribution, and asymptotic approximation may not be valid given the sample sizes common for this problem.

In our method, we attempt to overcome the weaknesses in existing methods described above. We do not convert the raw data into ranks or discretize them and only assume that the data distributions are symmetric about their location parameters. The null and alternative distributions will then be constructed via the empirical distribution, and the stopping boundaries obtained by bootstrap sampling from the symmetric empirical distributions. This approach is valid for any sample size and, as correlations between test statistics are automatically taken care of, the given nominal level can be approximately achieved.

3. SET-UP AND NOTATION FOR THE PROPOSED METHOD

For concreteness, we focus on the case of a two-stage GS design with primary and secondary endpoints. The general case of multi-stage designs with multiple endpoints is similar and will be briefly discussed. Let $(x_1,\ldots,x_n)$ be the responses of the primary endpoint; $(y_1,\ldots,y_n)$be those for the secondary endpoints. Here $x_1,\ldots,x_n$ are i.i.d. with X, with $E(X) = u_1$; $y_1,\ldots,y_n$ are i.i.d. with Y, with $E(Y) = u_2$. The stage-one interim analysis is for the first n_1 observations (typically $n_1 \approx n/2$, and for a k-stage trial, the sample size for each stage j is roughly $n_j \approx n/k$). The goal is to test the hypotheses $H_1 : u_1 = 0 \; vs \; u_1 > 0 \; and \; H_2 : u_2 = 0 \; vs \; u_2 > 0$, The hypothesis H_2 will be tested if and only if is rejected either at the first stage or at the second stage, *i.e*. the primary endpoint acts as the gatekeeper for the secondary endpoint. For this, standard test statistics for the primary and secondary endpoints at the two stages are constructed. For simplicity of exposition we assume they are of the form

$$T_1 = \sqrt{n_1}\,\frac{1}{n_1\sigma_{t,1}}\sum_{i=1}^{n_1} x_i, \quad T_2 = \sqrt{n}\,\frac{1}{n\sigma_{t,2}}\sum_{i=1}^{n} x_i,$$

$$S_1 = \sqrt{n_1}\,\frac{1}{n_1\sigma_{s,1}}\sum_{i=1}^{n_1} y_i, \quad S_2 = \sqrt{n}\,\frac{1}{n\sigma_{s,2}}\sum_{i=1}^{n} y_i$$

where the σ' s are the corresponding standard deviations.

For this problem, the family-wise error rate (FWER) requirement is, for some pre-specified nominal level α,

$$FWER = P\left(Reject\ at\ least\ one\ true\ H_j\ (j=1,2)\right) \le \alpha. \quad (1)$$

Let (c_1, c_2) and (d_1, d_2) be the stopping boundaries for (T_1, T_2) and (S_1, S_2), in the sense described below. The GS procedure is as follows:

Stage 1. Take n_1 observations $x_1,\ldots,x_{n1}$ from the primary endpoint and test H_1 using T_1. If $T_1 \le c_1$ continue to Stage 2. If $T_1 > c_1$, reject H_1 and take observations $y_1,\ldots,y_{n1}$ from the secondary endpoint and test H_2 by S_1. If $S_1 > d_1$, reject H_2; otherwise accept H_2. In either case terminate the trial.

Stage 2. Take additional observations $x_{n1+1},...,x_n$ from the primary endpoint and test H_1 by T_2. If $T_2 \le c_2$, accept H_1 and terminate the trial. If $T_2 > c_2$, reject H_1 and take additional observations $y_{n1+1},...,y_n$ from the secondary endpoints and test H_2 by S_2. If $S_2 > d_2$, reject H_2; otherwise accept H_2. Terminate the trial.

Now the problem amounts to the choice of the stopping boundaries (c_1, c_2) and (d_1, d_2) as described below·

The Stopping Boundaries

The critical boundaries (c_1, c_2) and (d_1, d_2) should be chosen to satisfy the FWER requirement (1). For this, the following two conditions (2) and (3) are to be satisfied.

$$FWER_1 = P_{H1}(T_1 > c_1) + P_{H1}(T_1 \le c_1, T_2 > c_2) \le \alpha. \quad (2)$$

$$FWER = P_{H2}(T_1 > c_1, S_1 > d_1) + P_{H2}(T_1 \le c_1, T_2 > c_2, S_2 > d_2) \le \alpha. \quad (3)$$

Any pair (c_1, c_2) satisfying (2) is referred to as *α*-level boundary for the primary endpoint. They may not be unique, *i.e*. there may be some other $(c_1', c_2') \ne (c_1, c_2)$ and (c_1', c_2') also satisfies (2). Condition (3) defines the joint boundaries (c_1, c_2) and (d_1, d_2) for the primary and secondary endpoints. Again, these boundaries are not unique and there are different commonly adopted ones. The O'Brien-Fleming boundary has the form $c_2 = \gamma c_1$, while the Pocock boundary are $c_2 = c_1$. More generally, we can use the error spending function of [13] to calculate the primary boundary. Let $\alpha(\cdot)$ be a non-decreasing function on [0, 1] with $\alpha(0) = 0$ and $\alpha(1) = \alpha$, then (c_1, c_2) can be calculated by

$$P_{H1}(T_1 > c_1) = \alpha\left(\frac{n_1}{n}\right), \text{ and } P_{H1}(T_1 > c_1) + P(T_1 \le c_1, T_2 > c_2) = \alpha(1) = \alpha.$$

4. THE PROPOSED METHOD

We concentrate on the case of two stages with primary and secondary endpoints. The case of multiple stage with multiple endpoints will be similar and will be briefly described at the end of the article. Although the case of multiple secondary endpoints was discussed in [3, 26, 12] discussed the case of multiple primary endpoints with Bonferroni correction of the nominal level. These methods do not consider the FWER requirement.

Let $F_1(.)$ be the distribution function of X, $F_2(.)$ be that of Y, and $F(.,.)$ be the joint distribution function of (X, Y). We construct the empirical distribution functions based on the observed data under assumption (A): the data distribution $F(.,.)$ is symmetric about (u_1, u_2). Given this assumption, we then obtain the empirical null and alternative distributions with which the stopping boundaries are derived. Calculation of statistical power follows in a straightforward manner. We note that assumption (A) implies the $F_j(.)$ are symmetric about $u_j\,(j = 1,2)$. It is a reasonable assumption for this problem, and much more flexible than assuming joint Normality. More generally, we can implement the location parameter (u_1, u_2) into the empirical distribution by empirical weights without any shape assumptions via the method of empirical likelihood. A drawback of this approach is that the empirical weights cannot be computed in closed form in general, and the effect of the parameters on the resulting empirical distribution is not directly visible.

By assumption (A) we have

$$F_1(x) = 1 - F_1(2\mu_1 - x - 0) \quad \textit{for all } x,$$

$$\textit{and } P((X,Y) \le (x,y)) = P((X,Y) \ge (2\mu_1 - x, 2\mu_2 - y)) \textit{ for all } (x,y). \textit{ Since}$$

$$P((X,Y) \ge (2\mu_1 - x, 2\mu_2 - y)) = 1 - P((X,Y) < (2\mu_1 - x, 2\mu_2 - y))$$

$$-P(X < 2\mu_1 - x, Y \ge 2\mu_2 - y) - P(X \ge 2\mu_1 - x, Y < 2\mu_2 - y)$$

$$= 1 - P((X,Y) < (2\mu_1 - x, 2\mu_2 - y)) - (P(X < 2\mu_1 - x) - P((X,Y) < (2\mu_1 - x, 2\mu_2 - y)))$$

$$-(P(Y < 2\mu_2 - y) - P((X,Y) < (2\mu_1 - x, 2\mu_2 - y)))$$

$$= 1 - F_1(2\mu_1 - x - 0) - F_2(2\mu_2 - y - 0) + F(2\mu_1 - x - 0, 2\mu_2 - y - 0)$$

Thus, under assumption (A) together with the above we have

$$F(x,y) = 1 - F_1(2\mu_1 - x - 0) - F_2(2\mu_2 - y - 0) + F(2\mu_1 - x - 0, 2\mu_2 - y - 0) \textit{ for all } (x,y)$$

Let $F_{1,n}(\cdot)$ be the empirical distribution function of $F_1(\cdot)$ based on the data $x_1,...,x_n$, $F_{2,n}(\cdot)$ be that of $F_2(\cdot)$ based on the data $y_1,...,y_n$, and $F_n(\cdot,\cdot)$ be the empirical distribution function of $F(\cdot,\cdot)$ based on the data $(x_1,y_1),....,(x_n,y_n)$. Then under assumption (A), for given μ_1, the empirical estimate $\hat{F}_{1,n}(\cdot \mid \mu_1)$ of F_1, F based on the data $x_1,...,x_n$ is (c.f. [10, 13]).

$$\hat{F}_{1,n}(x/\mu_1)=\frac{1}{2}\left(F_{1,n}(x)+1-F_{1,n}(2\mu_2-x-0)\right).$$

In fact, $\hat{F}_{1,n}(\cdot\,|\,\mu_1)$ is just the uniform distribution over the 2*n* points $\{x_1,...,x_n,2\mu_1-x_1,...,2\mu_1-x_n\}$. Similarly, for given (μ_1,μ_2), under assumption (A), the empirical estimate $\hat{F}_1(\cdot,\cdot\,|\,\mu_1,\mu_2)$ of $F(\cdot,\cdot)$ based on the data $(x_1,y_1),.....,(x_n,y_n)$ is

$$\hat{F}_n(x,y\,|\,\mu_1,\mu_2)=\frac{1}{2}(F_n(x,y)+1-F_{1,n}(2\mu_1-x-0)$$
$$-F_{2,n}(2\mu_2-y-0)+F_n(2\mu_1-x-0,2\mu_2-y-0))$$

and $\hat{F}_n(\cdot,\cdot\,|\,\mu_1,\mu_2)$ is just the uniform distribution over the 2*n* points $\{(x_1,y_1),...,(x_n,y_n),(2\mu_1-x_1,2\mu_2-y_1),...,(2\mu_1-x_n,2\mu_2-y_n)\}$.

Below we compute the stopping boundaries (c_1,c_2) = $(c_{n1,1},c_{n,2})$ and (d_1,d_2) = $(d_{n1,1},d_{n,2})$ under the empirical distributions $\hat{F}_{1,n1}(\cdot\,|\,0)$ and $\hat{F}_n(\cdot,\cdot\,|\,0,0)$, by bootstrap simulation described below. Let *N* be the total number of bootstrap samples (typically, $N \geq 10{,}000$), and we choose the error spending function as $\alpha(n_1/n)=\alpha n_1/n$.

Choice of $c_1=c_{n1,1}$. Draw m bootstrap samples $x_1^{(k)},...,x_{n1}^{(k)}$ from $\hat{F}_{1,n1}(\cdot\,|\,0)(k=1,...,N)$ under H_1, *i.e.* with $\mu_1=0$. For each sample k, compute the statistics $T_1^{(k)}(k=1,...,N)$, and set $c_{n1,1}$ be the solution of the equation

$$P_{H1}(T_1>c_1)=\alpha\left(\frac{n_1}{n}\right)$$

Then $c_{n1,1}$ is the $\left(1-\alpha\left(\frac{n_1}{n}\right)\right)$-th upper sample quantile of $T_1^{(1)},...,T_1^{(N)}$, which is the empirical $\left(1-\alpha\left(\frac{n_1}{n}\right)\right)$-th upper quantile of T_1 under the null hypothesis H_1.

Choice of $c_2=c_{n,2}$. For c_1 given in the above relation, we solve c_2 from the following equation, corresponding to expression (2),

$$\alpha=P_{H1}(T_1>c_1)+P_{H1}(T_1\leq c_1,T_2>c_2)$$
$$=P_{H1}(T_1>c_1)+P_{H1}(T_2>c_2\,|\,T_1\leq c_1)\left(1-P_{H1}(T_1>c_1)\right),$$

or, with the notation "$A:=B$" means denote A as B,

$$P_{H1}(T_2>c_2\,|\,T_1\leq c_1)=\frac{\alpha-P_{H1}(T_1>c_1)}{1-P_{H1}(T_1>c_1)}=\frac{\alpha-\alpha\left(\frac{n_1}{n}\right)}{1-\alpha\left(\frac{n_1}{n}\right)}:=\alpha=\alpha_{n,n_1}(\alpha)$$

i.e., c_2 is the $(1-\alpha)-th$ upper quantile of the conditional distribution $P_{H1}(T_2>c_2\,|\,T_1\leq c_1)$. Since the event $\{T_1\leq c_1\}$ has probability $1-\alpha(n_1/n)$, we draw $N_1=\left[N/\left(1-\alpha(n_1/n)\right)\right]$ (note $N_1\neq m$ in the above) bootstrap samples $x_1^{(k)},...,x_n^{(k)}$ from $\hat{F}_{1,n}(\cdot\,|\,0)(k=1,...,N_1)$ under H_1, *i.e.* with $\mu_1=0$ so that there are about N samples which satisfy the condition $T_1^{(k)}\leq c_1$, here $T_1^{(k)}$ is computed from the first n_1 components of $x_1^{(k)},...,x_n^{(k)}$. Without confusion denote $x_1^{(k)},...,x_n^{(k)}$, for $\hat{F}_{1,n}(\cdot\,|\,0)(k=1,...,N)$ be all such conditional samples Based on these N conditional samples, we compute $T_2^{(k)}(k=1,...,N)$ and set c_2 to be the $(1-\alpha)-th$ sample upper quantile of $T_2^{(1)},...,T_2^{(N)}$.

Choice of the Secondary Boundary

[31] Show that if there is a group sequential procedure (GSP) which tests every intersection hypothesis at level α, then by the closure principle of [20], the FWER of the GSP is of level α. For two endpoints, with H_1 and H_2, there are only three intersections, $H_1\cap H_2$, H_1 and H_2, and a level α test for $H_1\cap H_2$ is also level α for H_1 by the hierarchical testing of H_1 and H_2. However, given (c_1, c_2) there is no unique solution for (d_1, d_2) in equation (3). [11] studied the choice $d_1 = d_2 = Z_\alpha$, the upper $(1-\alpha)$-th quantile of the N (0,1) distribution. [29] show that this choice does not control the FWER for all correlations between the two endpoints with given μ_1. They also investigated three other configurations, $(c_1, c_2) = (d_1, d_2)$, $c_1 > d_1$ and $c_2 < d_2$, and $c_1 < d_1$ and $c_2 > d_2$, and concluded in their Propositions 2-4 that the level α can be achieved for some choices of (μ_1, μ_2) only when the correlation between the primary and secondary endpoints is one. Their determination of (d_1, d_2) cannot be given in closed form, as they are the solution of a non-linear integration equation.

We propose that it is more reasonable to choose $d_1 = d_{n1,1}$ to match c_1 for T_1. Since

$P_{H1}(T_1 > c_1) = \alpha(n_1/n)$, we choose d_1 such that

$$P_{H_2}(S_1>d_1)=\alpha(n_1/n).$$

Now, with (c_1, c_2, d_1) given above, we set $d_2 = d_{n,2}$ as the solution in d_2 of the following variant of equation (3)

$$\alpha = FWER = P_{H2}\left(T_1 > c_1, S_1 > d_1\right) + P_{H2}\left(T_1 \le c_1, T_2 > c_2, S_2 > d_2\right). \quad (4)$$

Remark. i) It is apparent that there is d_2 to achieve (4) if and only if

$$P_{H2}\left(T_1 > c_1, S_1 > d_1\right) + P_{H2}\left(T_1 \le c_1, T_2 > c_2\right) \ge \alpha. \quad (5)$$

This condition is not necessarily met for all data distributions, neither for the Normal approximation method. When $P_{H2}(T_1 > c_1, S_1 > d_1) + P_{H2}(T_1 \le c_1, T_2 > c_2) = \alpha$, (4) is satisfied with d_2 being the minimum point of support of S_2;

ii) When $P_{H2}(T_1 > c_1, S_1 > d_1) + P_{H2}(T_1 \le c_1, T_2 > c_2) < \alpha$, equality in (4) cannot be achieved. We choose d_2 such that

$$P_{H2}(S_1 > d_1) + P_{H2}(S_2 > d_2) = \alpha \quad (6)$$

and the actual FWER will be computed via simulation.

When condition (5) is satisfied, then

$$\alpha = P_{H_2}\left(T_1 > c_1, S_1 > d_1\right) + P_{H_2}\left(S_2 \mid > d_2 \mid T_1 \le c_1, T_2 > c_2\right) P_{H_2}\left(T_1 \le c_1, T_2 > c_2\right)$$

or

$$P_{H_2}\left(S_2 > d_2 \mid T_1 \le c_1, T_2 > c_2\right) = \frac{\alpha - P_{H_2}\left(T_1 > c_1, S_1 > d_1\right)}{P_{H_2}\left(T_1 \le c_1, T_2 > c_2\right)} := \frac{\alpha = p_1}{p_2} := \delta_2,$$

i.e., d_2 is the $(1 - \delta_2)$-th upper quantile of the conditional distribution $P_{H2}(S_2 > d_2 | T_1 \le c_1, T_2 > c_2)$. To determine d_2 we need first to compute the quantities p_1 and p_2.

Note that $p_1 = \int P_{H_2}\left(T_1 > c_1, S_1 > d_1 \mid \mu_1\right) dQ\left(\mu_1\right)$, where $Q(\cdot)$ is the distribution function of μ_1. As $Q(\cdot)$ is unknown, we estimate it by $\hat{Q}_n(\cdot)$ as below. For this, draw for $x_1^{(k)}, ..., x_n^{(k)}$ from the original marginal empirical distribution $F_{1,n}(\cdot)\left(k = 1, ..., N\right)$, and compute $\mu_n^{(k)} = n^{-1}\sum_{i=1}^{n} x_i^{(k)}$ $\left(k = 1, ..., N\right)$, Let $\hat{Q}_n(\cdot)$ be the empirical distribution function from $\mu_n^{(1)}, ..., \mu_n^{(N)}$. Then, for $k = 1, ..., N$, sample $\mu_1^{(k)} \sim \hat{Q}_n(\cdot)$, given this $\mu_1^{(k)}$ sample $\left(x_1^{(k)}, y_1^{(k)}\right), ..., \left(x_n^{(k)}, y_n^{(k)}\right) \sim \hat{F}_n\left(\cdot, \cdot \mid \mu_1^{(k)}, 0\right)$ Compute $T_1^{(k)}$ using the samples $x_1^{(k)}, ..., x_{n1}^{(k)}\left(k = 1, ..., N\right)$, compute $S_1^{(k)}$ using the samples $y_1^{(k)}, ..., y_{n1}^{(k)}\left(k = 1, ..., N\right)$, and compute $T_2^{(k)}$ using the samples $x_1^{(k)}, ..., x_n^{(k)}\left(k = 1, ..., N\right)$. Then p_1 and p_2 are well approximated by

$$\hat{p}_1 = \frac{1}{N}\sum_{k=1}^{N} I\left(T_1^{(k)} > c_1, S_1^{(k)} \mid > d_1\right),$$

$$\hat{p}_2 = \frac{1}{N}\sum_{k=1}^{N} I\left(T_1^{(k)} \le c_1, T_1^{(k)} > c_2\right),$$

where $I(\cdot,\cdot)$ is the indicator function. Now with $\hat{p}_1$ and $\hat{p}_2$, we set $\hat{\delta}_2 = \left(\alpha - \hat{p}_1\right)/\hat{p}_2$ as an estimate of δ_2. Lastly, we set d_2 as the $\left(1 - \hat{\delta}_2\right)$-th upper quantile of the conditional distribution $P_{H_2}\left(S_2 > d_2 \mid T_1 \le c_1, T_2 > c_2\right)$. For this using the bootstrap samples above, compute $S_2^{(k)}$ using the data $\left(y_1^{(k)}, ..., y_n^{(k)}\right)$ *for* $k = 1, ..., N$. Set $\tilde{N} = \sum_{k=1}^{N} I\left(T_1^{(k)} \le c_1, T_2^{(k)} > c_2\right)$, and $\tilde{S}_2^{(k)} = S_2^{(k)} I\left(T_1^{(k)} \le c_1, T_2^{(k)} > c_2\right)$,and set d_2 be the $\left(1 - \hat{\delta}_2\right)$-th upper sample quantile of the $\tilde{N}$ non-zero $\tilde{S}_2^{(k)}$'s.

From the above procedures, we see that in our method the FWER can achieve the pre-specified nominal level α for some data, regardless the dependence relationship between the primary and secondary endpoints. The actual proportion of data to achieve the nominal level depends on the set up of the data.

Power. With given $\mu_1 > 0$ and/or $\mu_2 > 0$, for given level α, we compute the primary power

$$\beta_1 = \beta_1\left(\alpha, \mu_1\right) = P_{\mu 1}\left(T_1 > \mid c_1\right) + P_{\mu 1}\left(T_1 \le c_1, T_2 > c_2\right), \quad (7)$$

and the secondary power

$$\beta_2 = \beta_2\left(\alpha, \mu_2\right) = P_{\mu 2}\left(T_1 > c_1, S_1 > d_1\right) + P_{\mu 2}\left(T_1 \le c_1, T_2 > c_2, S_2 > d_2\right). \quad (8)$$

To compute $\beta_1\left(\alpha, \mu_1\right)$, for $k = 1, ..., N$, sample $x_1^{(k)}, ..., n_n^{(k)}$, from $\hat{F}_{1,n}\left(\cdot \mid \mu_1\right)$, and compute $T_1^{(k)}$ and $T_2^{(k)}$, and β_1 is approximated by

$$\hat{\beta}_1 = \hat{\beta}_1\left(\alpha, \mu_1\right) = \frac{1}{N}\sum_{k=1}^{N}\left(I\left(T_1^{(k)} > c_1\right) + I\left(T_1^{(k)} \le c_1, T_2^{(k)} > c_2\right)\right).$$

To compute $\beta_2\left(\alpha, \mu_2\right)$, for $k = 1, ..., N$, sample $\mu_1^{(k)} \sim \hat{Q}_n$, where $\hat{Q}_n$ is given in the computation of d_2; then sample $\left(x_1^{(k)}, y_1^{(k)}\right), ..., \left(x_n^{(k)}, y_n^{(k)}\right) \sim \hat{F}_n\left(\cdot, \cdot \mid \mu_1^{(k)}, \mu_2\right)$,

and compute $T_1^{(k)}, S_1^{(k)}, T_2^{(k)}$ and $S_2^{(k)}$ and β_2 is estimated by

$$\hat{\beta}_2 = \hat{\beta}_2(\alpha, \mu_2) = \frac{1}{N}\sum_{k=1}^{N}\left(I\left(T_1^{(k)} > c_1, S_1^{(k)} > d_1\right) + I\left(T_1^{(k)} \le c_1, T_2^{(k)} > c_2, S_2^{(k)} > d_2\right)\right).$$

The case of two-arm and multi-stage with multiple endpoints is similar and not shown to save space, but is available upon request from the corresponding author.

5. SIMULATION STUDY

We consider the case of two stage one-armed case. Recall OF stands for the [22] boundaries, PO for [25] boundaries. The notation OF1-PO2 stands for the combination of [22] primary boundaries and Pocock's secondary boundaries; AH2 stands for the ad-hoc secondary boundaries. In the existing methods which use Normal approximations, the boundaries are fixed in advance and are independent of the observed data and study sample size. They depend on the correlation ρ between the primary and secondary endpoints, which is often unknown, and is often supplied by a guessed value. The boundaries also depend on the choice of which combination of methods are used.

This may perform poorly with small sample sizes. For example, with α=0.05, n_1=n, ρ=0.4, if OF1-PO2 combination is used, then $(c_1, c_2, d_1, d_2) = (1.678\sqrt{2}, 1.678, 1.686, 1.686)$; with the OF1-AH2 combination, $(c_1, c_2, d_1, d_2) = (1.678\sqrt{2}, 1.678, 1.714, 1.645)$; When n_1/n = 0.75, with α = 0.025, for Strategy 1 as in Tamhane *et al*. (2010) with (c_1, c_2, d_1, d_2) = (2.340, 2.012, 2.340, 2.012); for Strategy 2, (c_1, c_2, d_1, d_2) = (2.340, 2.012, 2.040, 2.258).

Below we compute stopping boundaries with the proposed method. We give three examples. The first example is significant for both the primary and secondary endpoints; the second is non-significant for both endpoints; and the third is significant for the primary endpoint but not for the secondary endpoint. With each example we compute the boundaries for a range of sample sizes, and for each sample size we sample three sets of data. The exact FWER is given in all cases. In all the three examples we set the error spending function as $\alpha(n_1/n) = \alpha n_1/n$.

Example 1. The total sample size is n, with first stage at n_1. $(x_i, y_i) \sim 0.5\varphi_1 + 0.5\varphi_2 \; (i = 1, \ldots n)$, with φ_1 be the normal density with mean (0,0) and covariance matrix $(1, 0.5; 0.5, 1)$, and φ_2 is the Normal density with mean $(1, 1.6)$ and covariance matrix $(1, -0.2, -0.2, 1)$. *i.e*. the data is from a two-component normal mixture which is symmetric around (μ_1, μ_2) = $(0.5, 0.8)$, but is not normal. Thus this example has significant deviation from both H_1 and H_2, and T_1, T_2, S_1 and S_2 are not normal.

The densities of the primary and secondary endpoints data of one sample draw from the above setting are plotted in Figure **1** below. Table **1** gives results (α = 0.05) of the boundaries computed from the

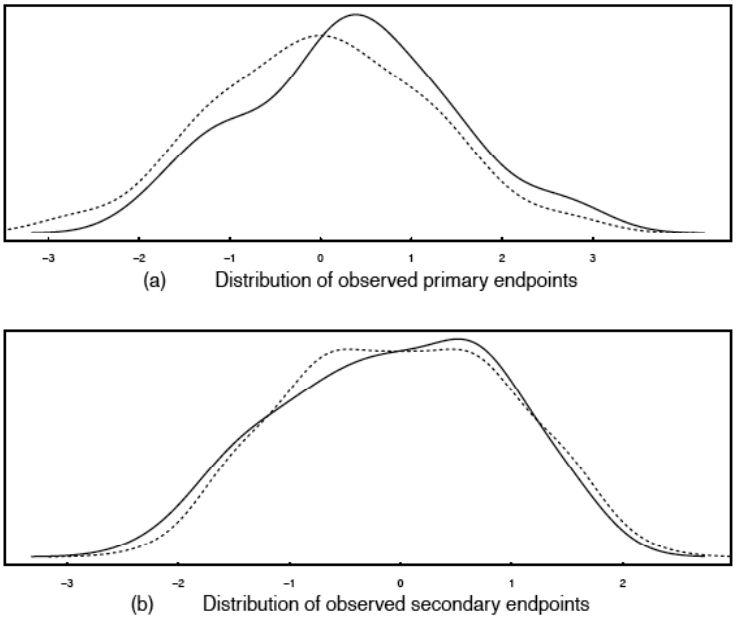

Figure 1: Densities for primary (**a**) and secondary (**b**) endpoints: solid line – original data; dashed line – symmetrized data.

Table 1: Boundaries with Proposed Method for Simulated Data

(n_1, n)	T_1	T_2	S_1	S_2	c_1	c_2	d_1	d_2	$FWER$
(10, 20)	2.687	2.411	2.714	2.779	1.893	1.605	1.865	0.823	0.05
	0.288	1.044	1.265	2.602	1.198	1.354	1.471	0.317	0.05
	1.665	1.416	1.531	2.439	1.496	1.416	1.531	0.994	0.05
(15, 35)	1.825	1.452	3.397	5.042	1.467	1.345	1.848	0.808	0.05
	1.573	2.345	1.437	3.395	1.499	1.396	1.484	1.025	0.05
	0.809	2.551	1.543	2.650	1.500	1.428	1.596	0.903	0.05
(20 50)	1.816	3.715	2.051	4.397	1.585	1.473	1.614	1.013	0.05
	1.571	3.521	4.293	5.442	1.554	1.483	1.787	0.558	0.05
	3.375	4.722	2.211	4.867	1.830	1.563	1.540	0.879	0.05
(40, 100)	3.237	4.902	6.777	7.159	1.665	1.466	2.034	0.368	0.05
	3.262	3.942	4.788	5.259	1.568	1.509	1.794	0.854	0.05
	3.228	4.107	4.278	6.233	1.620	1.364	1.678	0.491	0.05
(200, 500)	6.472	9.029	8.150	12.975	1.616	1.412	1.689	1.539	0.02
	6.312	10.822	9.994	14.895	1.593	1.457	1.654	1.559	0.02
	7.039	10.669	8.612	14.195	1.643	1.467	1.661	1.561	0.02
(500, 1000)	8.379	12.163	13.041	19.419	1.505	1.460	1.607	1.610	0.025
	9.597	14.584	13.222	19.535	1.470	1.531	1.620	1.631	0.025
	10.350	14.170	13.250	19.255	1.540	1.531	1.582	1.590	0.025

proposed method. As the results differ from different draws under the same setting, we display the results for different sample sizes, with three samples for each fixed sample sizes (n_1, n).

We see from Table **1** that, for the proposed method, these quantities are highly dependent on the observed data and the sample sizes. The estimates are therefore more reasonable when the sample sizes are not large and the data distribution deviates from the Normal. As the data are not Normal, the boundaries in Table **1** tend to be smaller than those given by Normal approximation. When the nominal level α is achieved, the corresponding d_2 is small. In some cases the nominal level $(\alpha = 0.05)$ is not achieved, and the corresponding d_2 is bigger. In these cases the boundaries are conservative, as the boundaries from Normal approximations. All these results rejected H_1 and H_2 based on the simulated boundaries.

Example 2. We sample $(x_i, y_i) \sim 0.5\phi_1 + 0.5\phi_2 \, (i = 1, ...n)$, with ϕ_1 be the Normal density with mean $(-0.6, -0.5)$ and covariance matrix $(1, 0.5; 0.5, 1)$, and ϕ_2 is the Normal density with mean $(0.6, 0.5)$ and covariance matrix $(1, -0.2, -0.2, 1)$, *i.e*. The data is from a two-component Normal mixture which is symmetric around $(\mu_1, \mu_2) = (0, 0)$. Thus this example has no deviation from both H_1 and H_2. In this example we set $\alpha = 0.025$. The results are shown in Table **2**.

We see that for the first data with sample size $(15, 35)$ that, $S_2 = 1.328 > d_2 = 0.557$, and hence H_2 is rejected. The other results are correctly accepted both H_1 and H_2. We see that in most of these cases, the FWER of 0.025 is achieved, and in many cases the d_2's are small and hence the acceptance of H_2 is quite stringent.

Example 3. We sample $(x_i, y_i) \sim 0.5\phi_1 + 0.5\phi_2 \, (i = 1, ...n)$, with ϕ_1 be the Normal density with mean $(0, -0.5)$ and covariance matrix $(1, 0.5; 0.5, 1)$, and ϕ_2 is the Normal density with mean $(0.6, 0.5)$ and covariance matrix $(1, -0.2, -0.2, 1)$, *i.e*. the data is from a two-component Normal mixture, which is symmetric around $(\mu_1, \mu_2) = (0.6, 0)$. Thus, this example has no-true H_1 and true H_2. In this example we set $\alpha = 0.025$.

The results are shown in Table **3**.

Two incorrect conclusions were observed in these simulations. For the second set of data at sample size (10,20), both H_1 and H_2 are rejected due to large values of T_j and S_j, although in truth should only rejected H_1.

Table 2: Boundaries with Proposed Method for Simulated Data

(n_1, n)	T_1	T_2	S_1	S_2	c_1	c_2	d_1	d_2	$FWER$
(10, 20)	1.374	0.844	0.757	0.174	1.723	1.617	1.475	0.672	0.025
	-1.060	-0.639	-0.877	-0.503	1.573	1.461	1.521	-0.214	0.025
	-0.100	-0.205	-1.500	-1.499	1.436	1.526	1.564	0.284	0.025
(15, 35)	0.234	-0.357	1.194	1.328	1.453	1.579	1.581	0.557	0.025
	1.164	1.657	0.437	-0.548	1.546	1.717	1.433	0.992	0.025
	-1.503	-2.584	-0.732	-0.359	1.631	1.842	1.549	1.703	0.015
(20 50)	-0.082	0.772	-0.879	-1.441	1.404	1.462	1.477	0.618	0.025
	-2.249	-2.496	-0.567	-0.641	1.955	1.615	1.628	1.428	0.01
	0.180	-0.729	-1.275	-0.226	1.496	1.511	1.639	0.045	0.025
(40, 100)	-0.553	0.284	0.991	0.302	1.507	1.471	1.612	0.302	0.025
	-1.282	-0.964	-1.270	-0.525	1.660	1.605	1.726	1.774	0.001
	-0.590	-0.060	-0.625	-0.416	1.543	1.437	1.736	0.372	0.025
(200, 500)	-0.340	-0.100	-1.286	-1.475	1.586	1.544	1.714	0.341	0.025
	1.441	0.976	0.791	0.033	1.581	1.451	1.591	0.849	0.025
	0.248	-0.139	-0.835	-1.271	1.650	1.514	1.652	0.358	0.025
(500, 1000)	0.439	0.272	0.937	0.213	1.651	1.611	1.639	0.498	0.025
	-0.028	-1.059	-0.801	-0.686	1.596	1.527	1.582	-0.461	0.025
	0.248	0.633	-0.427	1.226	1.581	1.640	1.568	0.719	0.025

Table 3: Boundaries with Proposed Method for Simulated Data

(n_1, n)	T_1	T_2	S_1	S_2	c_1	c_2	d_1	d_2	$FWER$
(10, 20)	0.829	2.004	0.317	-1.174	1.422	1.538	1.362	0.672	0.025
	1.429	2.797	2.634	3.136	1.694	1.757	2.098	1.562	0.025
	1.256	1.820	1.250	1.045	1.785	1.689	1.459	0.957	0.025
(15, 35)	0.581	1.278	-1.400	-0.389	1.551	1.722	1.528	0.885	0.025
	0.843	1.794	-0.909	-0.350	1.626	1.573	1.574	0.987	0.025
	4.1743	3.462	0.400	0.353	2.386	1.699	1.621	1.176	0.015
(20 50)	2.291	2.105	-1.293	-1.149	1.822	1.507	1.585	1.258	0.025
	0.578	2.081	-0.869	-1.076	1.574	1.617	1.643	1.187	0.01
	0.365	2.047	-0.241	0.322	1.544	1.635	1.554	1.015	0.025
(40, 100)	3.131	3.783	-2.224	-1.758	1.749	1.446	1.615	0.822	0.025
	2.134	2.865	-0.418	-0.176	1.620	1.546	1.615	1.074	0.001
	2.914	4.085	-0.360	-0.190	1.675	1.579	1.592	1.114	0.025
(200, 500)	2.890	6.172	1.067	0.961	1.641	1.560	1.682	1.447	0.010
	4.106	5.910	0.874	-0.286	1.633	1.576	1.688	1.551	0.010
	2.882	5.047	-0.595	-2.702	1.676	1.553	1.633	0.276	0.025
(500, 1000)	7.094	8.335	-0.407	-0.395	1.649	1.562	1.698	1.555	0.013
	6.617	9.673	0.622	0.697	1.641	1.678	1.538	1.594	0.013
	5.854	8.138	-0.715	-1.502	1.560	1.623	1.505	1.521	0.013

For the first set of data with sample size $(15,35)$, H_1 is not rejected due to small value of T_2, which is not correct. All the other cases reached the correct decisions.

DISCUSSION

We investigated a robust method for determining stopping boundaries subject to a family wise error rate requirement for testing primary and secondary endpoints in clinical trials with small sample sizes. We considered the one-arm and two-arm cases, but only reported the one-arm case here. Our method assumes the data follows a symmetric distribution about their location parameters, and makes no assumptions on the distributions of the test statistics, and hence is robust. Its performance contrasts with existing methods for determining the boundaries. These use Normal approximations for the test statistics, their stopping boundaries are determined in advance before observing the data, and do not achieve the FWER except for a few special cases. With the proposed method, these boundaries are determined according to the observed data, and FWER is achieved in many cases. The proposed method was illustrated with three simulation examples for the one-armed case.

To relax the assumption of symmetric distribution of the data, it is possible to use the empirical likelihood method to implement the location parameter. This will improve robustness of the method further but at the cost of not giving a closed form for the corresponding construction of the weighted empirical distributions. This approach also relies on large sample approximations, which makes it unsuitable for our objective of providing a method for small sample size studies.

The proposed method is mainly for continuous observations or discrete observations with relatively large number of possible values. The symmetry assumption is not necessary when both primary and secondary endpoints are discrete and have a small number of possible values. In this case, the small number of parameters are easily estimated by MLE, and exact inference can still be performed. For example, consider the case where primary and secondary endpoints are binary variables. Let $p = P(X=1) = 1 - P(X=0)$ *and* $q = P(Y=1) = 1 - P(Y=0)$. In this case, the hypotheses are $H_1 : p = p_0$ *vs* $H_1' : p > p_0$ *and* $H_2 : q \le q_0$ *vs* $H_2' : q > q_0$, for some given p_0 and q_0. The joint distribution of (X, Y) is characterized by the probabilities $p_{ij} = P(X=i, Y=j)(i, j = 0,1)$. The MLE $\hat{p}_{n1}, \hat{p}_n, \hat{q}_{n1}, \hat{q}_n, \hat{p}_{ij,n1}$ *and* $\hat{p}_{ij,n}$ at stages 1 and 2 are obtained in closed form, and $T_1 := \sum_{i=1}^{n1} x_i \approx Bino(n_1, \hat{p}_{n1}), S_1 := \sum_{i=1}^{n1} y_i \approx Bino(n_1, \hat{q}_{n1})$ similarly for T_2 and S_2. Then stopping boundaries (c_1, c_2, d_1, d_2) as solutions of expressions (2) and (3) can be obtained via bootstrap simulations. In particular, c_1 and c_1 are determined by (2), with c_1 first determined by the error spending function as given before. We can determine d_1 in the same way as c_1. To determine d_2 we need the joint distribution of (T_1, S_1) and (T_1, T_2, S_2), which can be obtained via bootstrap simulation with the estimated parameters $\hat{p}_{ij}$'s.

The method can also be modified for the case where one of the endpoints is continuous and the other is discrete with a small number of possible values. For example, suppose the secondary endpoint is binary with $q = P(Y=1) = 1 - P(Y=0)$. Let $\hat{q}_{n1}$ and $\hat{q}_n$ be the MLE of q at stages 1 and 2. In this case, the hypotheses are $H_1 : \mu = 0$ and $H_2 : q \le q_0$. Let $F_j(\cdot \mid \mu)$ be the distribution function of X given $Y = j(j = 0,1)$ and $F_j(\cdot \mid \mu)$ be the marginal distribution of X. We assume they are symmetric around μ, and the corresponding empirical distribution functions can be constructed. The boundaries (c_1, c_2, d_1, d_2) can be obtained via bootstrap by the way described in Section 3.1 and above.

REFERENCES

[1] 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; 61: 377-384. https://doi.org/10.1016/S0009-9236(97)90170-3

[2] Berry DA. Interim analysis in clinical trials: Classical vs. Bayesian approaches. Statistics in Medicine 1985; 4: 521526. https://doi.org/10.1002/sim.4780040412

[3] Davis C. Secondary endpoints can be validly analyzed, even if the primary endpoints do not provide clear statistical significance, Controlled Clinical Trials 1997; 18: 557-560. https://doi.org/10.1016/S0197-2456(96)00133-X

[4] Edgington ES. (1987). Randomization Tests, 2nd ed. New York, Marcel Dekker.

[5] Emerson SS, Fleming TR. Parameter estimation following sequential hypothesis testing, Biometrika 1990; 77: 875-892. https://doi.org/10.1093/biomet/77.4.875

[6] Fisher RA. On the interpretation of χ^2 from contingency tables, and the calculation of p. Journal of the Royal Statistical Society 1922; 85 (1): 8794. https://doi.org/10.2307/2340521

[7] Fisher RA. (1954). Statistical Methods for Research Workers. Oliver and Boyd.

[8] Good P. (1993). Permutation Tests, Springer-Verlag, New York.

[9] Graubard BI and Korn EL. Choice of column scores for testing independence in ordered 2 × K contingency tables, Biometrics 1987; 43: 471-476. https://doi.org/10.2307/2531828

[10] Hinkley D. On estimating a symmetric distribution, Biometrika 1976; 63: 680681. https://doi.org/10.1093/biomet/63.3.680

[11] Hung J, Wang SJ, O'Neill R. Statistical considerations for testing multiple endpoints in group sequential or adaptive clinical trials, Journal of Biopharmaceutical Statistics 2007; 17: 1201-1210. https://doi.org/10.1080/10543400701645405

[12] Jennison C, Turnbull B. (2000). Group Sequential Methods with Applications to Clinical Trials, Chapman and Hall: Boca Raton, FL.

[13] Lai TL, Robins H, Yu KF. Adaptive choice of mean or median in estimating the center of a symmetric distribution, Proc. Nati. Acad. Sci. USA 1983; 80: 5803-5806. https://doi.org/10.1073/pnas.80.18.5803

[14] Lan KKG, DeMets DL. Discrete sequential boundaries for clinical trials, Biometrika 1983; 70: 659-663. https://doi.org/10.2307/2336502

[15] Lo S. Estimation of a symmetric distribution, Annals of Statistics 1985; 13: 1097-1113. https://doi.org/10.1214/aos/1176349658

[16] Liu A, Hall WJ. Unbiased estimation following group sequential test, Biometrika 1999; 86: 71-78. https://doi.org/10.1093/biomet/86.1.71

[17] Liu A, Tan M, Boyett JM, Xiong X. Testing secondary hypothesis following sequential clinical trials, Biometrics 2000; 56: 640-644. https://doi.org/10.1111/j.0006-341X.2000.00640.x

[18] Mehta CR. and Patel NR. A network algorithm for performing Fisher's exact test in r × c contingency tables, Journal of the American Statistical Association 1983; 78: 427-434. https://doi.org/10.2307/2288652

[19] Mehta CR and Patel NR. (2010). IBM SPSS Exact Tests. SPSS Inc.

[20] Marcus R, Peritz E, Gabriel KR. On closed testing procedures with special references to ordered analysis of variance, Biometrika 1976; 63: 655-660. https://doi.org/10.1093/biomet/63.3.655

[21] Moss AJ, Hall WJ, Cannom 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; 335: 1933-1940. https://doi.org/10.1056/NEJM199612263352601

[22] O'Brien PC, Fleming TR. A multiple testing procedure for clinical trials, Biometrics 1979; 35: 549-556. https://doi.org/10.2307/2530245

[23] O'Neill PC. Secondary endpoints cannot be validly analyzed if the primary endpoints do not demonstrate clear statistical significance, Controlled Clinical Trials 1997; 18: 557-560. https://doi.org/10.1016/S0197-2456(97)00075-5

[24] Owen AB. Empirical likelihood for linear models, Annals of Statistics 1991; 19: 1725-1747. https://doi.org/10.1214/aos/1176348368

[25] Pocock SJ. Group sequential methods in the design and analysis of clinical trials, Biometrika 1977; 64: 191-199. https://doi.org/10.1093/biomet/64.2.191

[26] Prentice RL. On the role and analysis of secondary outcomes in clinical trials, Controlled Clinical Trials 1997; 18: 561-567. https://doi.org/10.1016/S0197-2456(96)00105-5

[27] Qin J and Lawless JL. Empirical likelihood and general estimating equations, Annals of Statistics 1994; 22: 300-325. https://doi.org/10.1214/aos/1176325370

[28] Senchaudhuri P, Mehta CR, Patel NR. Estimating exact p values by the method of control variables, or Monte Carlo rescue, Journal of the American Statistical Association 1995; 90: 640-648.

[29] Tamhane AC, Mehta CR, Liu L. Testing a primary and secondary endpoint in a group sequential design, Biometrics 2010; 66(4): 1174-1184. https://doi.org/10.1111/j.1541-0420.2010.01402.x

[30] Tan M, Xiong X and Kutner MH. Clinical trial designs based on sequential conditional probability ratio tests and reverse stochastic curtailing, Biometrics 1998; 54: 682-695. https://doi.org/10.2307/3109774

[31] Tang DI, Geller NL. Closed testing procedure for group sequential clinical trials with multiple endpoints, Biometrics 1999; 55: 1188-1192. https://doi.org/10.1111/j.0006-341X.1999.01188.x

[32] Wang SK and Tsaitis AA. Approximately optimal one-parameter boundaries for group sequential trials, Biometrics 1987; 43: 193-200. https://doi.org/10.2307/2531959

[33] Whithead J. On the maximum likelihood estimation following a sequential test, Biometrika 1986; 73: 573-581. https://doi.org/10.1093/biomet/73.3.573

[34] Xiong X, Tan M and Boyett J. Sequential conditional probability ratio test for normalized test statistics on information time, Biometrics 2003; 59: 624-631. https://doi.org/10.1111/1541-0420.00072

[35] Yates F. Test of significance for 2 × 2 contingency tables, Journal of the Royal Statistical Society 1984; A(75): 579.

[36] Ye Y, Li A, Liu L, Yao B. A group sequential Holm procedure with multiple primary endpoints, Statistics in Medicine 2012.

[37] Zhang J, Quan H, Ng J, Stepanavage M. Some Statistical Methods for Multiple Endpoints in Clinical Trials, Controlled Clinical Trials 1997; 18: 204-221. https://doi.org/10.1016/S0197-2456(96)00129-8

Received on 31-08-2017 Accepted on 10-10-2017 Published on 27-12-2017