Chapter 11
Critical Boundary Refinement in a Group Sequential Trial When the Primary Endpoint Data Accumulate Faster Than the Secondary Endpoint

Jiangtao Gou and Oliver Y. Chén

11.1 Introduction

In classical clinical trial studies, a clinical endpoint is defined as the time point at which a disease or symptom occurs. An individual reaching an endpoint during a clinical trial indicates either the conclusion of the trial, or there is strong evidence rendering the subject withdraws from the trial. To allow for early diagnosis, personalized treatment, and timely drug development, modern clinical trials are designed with customized endpoints. Consequently, the assessment time available to statistical analysis for each endpoint varies. For example, in oncology clinical trials, depending on the centering focus, the endpoints can be categorized into patient-centered endpoints and tumor-centered endpoints. An example of a patient-centered endpoint is the overall survival (OS), defined as the cumulative days a patient has lived, counting beginning from the date on which the disease is diagnosed or the date on which treatment is initiated; an example of a tumor-centered endpoint is progression-free survival (PFS), defined as cumulative days a patient has lived with cancer since the treatment and that the disease has not progressed (Fiteni et al. 2014). While OS is more reliable (since it covers a longer period) than PFS,

J. Gou (✉)
Department of Mathematics and Statistics, Hunter College of CUNY, New York, NY, USA

Department of Mathematics and Statistics, Villanova University, Villanova, PA, USA
e-mail: jgou@u.northwestern.edu

O. Y. Chén
Institute for Biomedical Engineering, University of Oxford, Oxford, UK
e-mail: yibing.chen@seh.ox.ac.uk

L. Zhang et al. (eds.), *Contemporary Biostatistics with Biopharmaceutical Applications*, ICSA Book Series in Statistics,
https://doi.org/10.1007/978-3-030-15310-6_11

the latter is usually used in practice as a surrogate for OS, when an accelerated evaluation is demanded, for example, during a drug test. However, we cannot at present ascribe such a replacement to any well-defined statistical theory, owning, in part, to the genuine differences between the two types of endpoints, and, in part, to the intellectual discovery of only modest correlation between PFS and OS (Amir et al. 2012; Michiels et al. 2017). By probing into the hierarchical basis of these two types of endpoints, statistical science can help us uncover the utility underlying each endpoint in addressing problems in clinical trials and improve statistical power. A beginning in this direction can be made by considering a hierarchical test embedded in a group sequentially design (Hung et al. 2007).

A group sequential design is a framework that allows statistical analysis during longitudinally ordered stages, defined as interim stages followed by a final stage (Jennison and Turnbull 2000). During each interim stage, a statistic (e.g. the estimated logarithm of the hazard ratio) is computed on data hitherto collected to determine whether or not to reject a null hypothesis (e.g. whether or not a treatment is more effective than the standard treatment), based upon a stopping criterion (called a critical boundary). Specifically, if the statistic exceeds the critical boundary, the null hypothesis is rejected, and the trial is subsequently terminated prior to the next interim stage. If a trial reaches the final stage, all data are utilized to test the null hypothesis.

Chief to a group sequential design is the critical boundary for early stopping. Pocock (1977) and O'Brien and Fleming (1979) individually proposed two now widely used critical boundaries for group sequential trials. Attributing to their contribution, these boundaries are commonly referred to as the Pocock (POC) boundary and the O'Brein-Fleming (OBF) boundary today, respectively. However, the POC and OBF boundaries require that the total number of decision times specified in advance. When this condition is not met, Lan and DeMets (1983) utilized a family of error spending functions to approximate the POC and the OBF boundaries. All of these approaches consider group sequential trials with a single primary endpoint. To address issues in group sequential trials involving multiple primary endpoints, Jennison and Turnbull (1993), Tang and Geller (1999), Maurer and Bretz (2013), Ye et al. (2013) and Xi and Tamhane (2015) provided various suggestions.

To raise any clinical finding related to an endpoint to the rank of science, one has to construct statistical hypotheses test for each endpoint. In a randomized trial consisting multiple endpoints, the endpoints often present a hierarchical structure. Statistical testing can be conducted serially for each ordered endpoint, or in parallel for all endpoints by applying the gatekeeping procedure (Dmitrienko and Tamhane 2007; Dmitrienko et al. 2009). A more flexible framework is the graph-theoretic-based procedure introduced by Bretz et al. (2009) and Burman et al. (2009), wherein nodes are used to represent hypothesis tests, coupled by directed and weighted edges indicating multiple test procedures. The above approaches were initially employed in single-stage designs with neither interim analysis nor trial extension. To extend these methods to multi-stage designs, Hung et al. (2007) first considered hierarchically testing multiple endpoints in a group sequential design.

The theoretical basis of group sequential designs involving multiple endpoints with complex hierarchical structure, one of the common practice in modern clinical trials, however, is not as-of-yet well-charted in statistical science. For instance, in an oncology trial, when the primary endpoint is PFS and the secondary endpoint is OS with the partially hierarchical design, can we improve upon the simple Bonferroni-based split between the primary and the secondary endpoint (which is the current practice), in a group sequential design? Prior work has built a reliable and useful repertoire that has offered us much insight, with which we build our theory. For example, Hung et al. (2007), Tamhane et al. (2010), Glimm et al. (2010), and Tamhane et al. (2018) considered the group sequential procedures for a primary and a secondary hypothesis with the same information fractions at interim analyses. In the light of their knowledge, in this article we attempt to address a few core issues in clinical trails when multiple objectives with hierarchical structures are present in group sequential designs.

11.2 Preliminaries

Consider a trial on a primary and a secondary endpoint hierarchically using a group sequential design with two stages. In the following, we use X to denote parameters and statistics that are related to the primary endpoint, and Y to denote parameters and statistics for the secondary endpoint. The number of interim looks at the secondary endpoint is permitted to be greater than the number of looks at the primary, if it takes longer to collect the secondary endpoint data than the primary endpoint data. We first consider a two-stage group sequential design that is applied to the primary endpoint, and a K-stage design that is used for the secondary endpoint ($K \geq 2$). For simplicity, we call it $[2|K]$-stage design. As a natural extension, we introduce the procedure with a K_X-stage design for the primary hypothesis and a K_Y-stage design for the secondary hypothesis. We denote this as a $[K_X|K_Y]$-stage design.

In a $[2|K]$-stage design, let $n_{1,X}$ and $n_{2,X}$ be the sample sizes for the two stages of the primary endpoint H_X, and $n_{1,Y}, n_{2,Y}, \ldots, n_{K,Y}$ for the K stages of the secondary endpoint H_Y. The total sample size is N, where $N = n_{1,X} + n_{2,X} = \sum_{i=1}^{K} n_{i,Y}$. The information time of the primary endpoint at the interim analysis is denoted as $t_X = n_{1,X}/N$. For the secondary endpoint, there are $K - 1$ interim analyses, and the information times are $t_{i,Y} = \sum_{j=1}^{i} n_{j,Y}/N$, $i = 1, \cdots, K - 1$. The information time or information fraction is the proportion of subjects or events already observed (Lan and DeMets 1989). The correlation between the two endpoints is denoted as ρ.

Let (X_1, X_2) and $(Y_1, Y_2, \ldots, Y_K)$ denote the standardized sample mean test statistics for the two endpoints at different stages, specified by a numeric subscript.

The normal theory applies asymptotically in this case. The correlations between the test statistics are shown as follows.

$$\operatorname{corr}(X_1, X_2) = \lambda, \quad \operatorname{corr}\left(Y_i, Y_j\right) = \gamma_i/\gamma_j \ (i < j),$$
$$\operatorname{corr}(X_1, Y_K) = \lambda\rho, \quad \operatorname{corr}(X_2, Y_K) = \rho,$$
$$\operatorname{corr}(X_1, Y_i) = \min\{\lambda/\gamma_i, \gamma_i/\lambda\} \cdot \rho, \quad \operatorname{corr}(X_2, Y_i) = \gamma_i\rho,$$

where $\lambda = \sqrt{t_X}$, $\gamma_i = \sqrt{t_{i,Y}}$ for $i = 1, \ldots, K-1$ and $\gamma_K = 1$.

Let $\left(\Delta_{1,X}, \Delta_{2,X}\right)$ and $\left(\Delta_{1,Y}, \Delta_{2,Y}, \ldots, \Delta_{K,Y}\right)$ denote the standardized treatment effects of the primary and the secondary endpoints at each stage. Noting that $\Delta_{1,X} = \lambda\Delta_{2,X}$ and $\Delta_{i,Y} = \gamma_i\Delta_{K,Y}$, we therefore simplify the notations by letting $\Delta_X = \Delta_{2,X}$ and $\Delta_Y = \Delta_{K,Y}$.

Denote H_X and H_Y as the primary and the secondary null hypotheses. Let (c_1, c_2) and $(d_1, d_2, \ldots, d_K)$ denote the primary boundary and the secondary boundary, respectively, in a group sequential procedure. Here, (c_1, c_2) correspond to (X_1, X_2) and $\left(\Delta_{1,X}, \Delta_{2,X}\right)$; $(d_1, d_2, \ldots, d_K)$ are with respect to $(Y_1, Y_2, \ldots, Y_k)$ and $\left(\Delta_{1,Y}, \Delta_{2,Y}, \ldots, \Delta_{K,Y}\right)$. Examples of common boundaries are discussed in Pocock (1977), O'Brien and Fleming (1979), and Lan and DeMets (1983).

In this article, we investigate three types of hierarchical testing scenarios: stage-wise hierarchical, overall hierarchical, and partially hierarchical scenarios. To conduct hypothesis testing with respect to each scenario, a scenario-specific decision rule needs to be defined a priori. Following Glimm et al. (2010), these decision rules are specified as below. Here, we define α_Y^S, α_Y^O, and α_Y^P, as the type I errors for a stagewise (S), an overall (O), and a partially (P) hierarchical rule, respectively, under the null hypothesis H_Y.

- Stagewise hierarchical rule $\mathcal{P}_S$. The primary hypothesis is tested sequentially. The secondary hypothesis will be automatically accepted if the primary hypothesis is not rejected. If the primary hypothesis is rejected, the secondary hypothesis will be tested only once at the same stage. The associated type I error is

$$\alpha_Y^S = \Pr(X_1 > c_1, Y_1 > d_1) + \Pr(X_1 \le c_1, X_2 > c_2, Y_2 > d_2).$$

- Overall hierarchical rule $\mathcal{P}_O$. Besides $\mathcal{P}_S$, the secondary hypothesis can be tested until its final stage if the primary hypothesis is rejected. The associated type I error is

$$\alpha_Y^O = \alpha_Y^S + \sum_{i=1}^{K-1} \Pr(X_1 > c_1, Y_1 \le d_1, \cdots, Y_i \le d_i, Y_{i+1} > d_{i+1})$$
$$+ \sum_{i=2}^{K-1} \Pr(X_1 \le c_1, X_2 > c_2, Y_2 \le d_2, \cdots, Y_i \le d_i, Y_{i+1} > d_{i+1}).$$

- Partially hierarchical rule $\mathcal{P}_P$. Besides $\mathcal{P}_O$, the secondary hypothesis can be tested from stage 2 to stage K if the primary hypothesis is failed to be rejected at its interim and final stage. The associated type I error is

$$\alpha_Y^P = \alpha_Y^O + \Pr\left(X_1 \le c_1, X_2 \le c_2, Y_2 > d_2\right) + \sum_{i=2}^{K-1} \Pr\left(X_1 \le c_1, X_2 \le c_2, Y_2 \le d_2, \cdots, Y_i \le d_i, Y_{i+1} > d_{i+1}\right).$$

Glimm et al. (2010) also listed another hierarchical rule called the coequal rule $\mathcal{P}_C$, where the primary and the secondary hypotheses are tested independently without any hierarchical structure. For a trial design using the coequal hierarchical rule, Bonferroni-type methods have been well developed, such as Maurer and Bretz (2013)'s method based on the graphical approach (Bretz et al. 2009, 2011), and Ye et al. (2013)'s method based on the Holm (1979) procedure. Other distribution-based or p-value-based tests can also be applied in trial designs using the coequal hierarchical rule, such as the Dunnett and Tamhane (1992) test, the Simes (1986) test, the generalized Simes test (Sarkar 2008; Gou and Tamhane 2014, 2018b), and their corresponding multiple testing procedures, such as Hommel (1988), Hochberg (1988), Rom (1990), and the hybrid Hochberg–Hommel procedure (Gou et al. 2014; Gou and Tamhane 2018a; Tamhane and Gou 2018). Since the endpoints under the coequal hierarchical rule are co-primary endpoints without a real hierarchical structure, we focus on the stagewise (S), the overall (O), and the partially (P) hierarchical rule in this article.

In a $[K_X|K_Y]$-stage design, we use terminologies and notations similar to those of a $[2|K]$-stage design. The sample sizes for H_X and H_Y in each stage are denoted as $n_{1,X}, \ldots, n_{K_X,X}$ and $n_{1,Y}, \ldots, n_{K_Y,Y}$ respectively, and the total sample size $N = \sum_{i=1}^{K_X} n_{i,X} = \sum_{i=1}^{K_Y} n_{i,Y}$. The cumulative sample sizes at stage i for H_X and H_Y are $N_{i,X} = \sum_{j=1}^{i} n_{j,X}$ and $N_{i,Y} = \sum_{j=1}^{i} n_{j,Y}$. The information times are calculated accordingly as $t_{i,X} = N_{i,X}/N$ and $t_{i,Y} = N_{i,Y}/N$, where $t_{K_X,X} = t_{K_Y,Y} = 1$. Let $\lambda_i = \sqrt{t_{i,X}}$, $\gamma_i = \sqrt{t_{i,Y}}$, and the correlation between X_{K_X} and Y_{K_Y} be ρ. The correlations between the standardized test statistics $(X_1, \cdots, X_{K_X})$ and $(Y_1, \cdots, X_{K_X})$ are

$$\text{corr}\left(X_i, X_j\right) = \lambda_i/\lambda_j \ (i < j), \quad \text{corr}\left(Y_i, Y_j\right) = \gamma_i/\gamma_j \ (i < j),$$
$$\text{corr}\left(X_i, Y_j\right) = \min\left\{\lambda_i/\gamma_j, \gamma_j/\lambda_i\right\} \cdot \rho, \quad \text{corr}\left(X_{K_X}, Y_{K_Y}\right) = \rho.$$

The standardized effects for H_X and H_Y at the final stage are denoted as Δ_X and Δ_Y, so the effects at interim stage i are $\lambda_i \Delta_X$ and $\gamma_i \Delta_Y$, respectively. The critical boundaries for standardized test statistics of H_X and H_Y are $(c_1, \cdots, c_{K_X})$ and $(d_1, \cdots, d_{K_Y})$. When $K_X = K_Y$, Tamhane et al. (2018) gave the expressions of type I error rates under H_Y for $\mathcal{P}_S$, $\mathcal{P}_O$, and $\mathcal{P}_P$. In a more general setting when $K_X \neq K_Y$, the corresponding type I error rates under H_Y are

$$\mathcal{P}_S : \alpha_Y^S = \sum_{i=1}^{K_X \wedge K_Y} \Pr\left(X_1 \leq c_1, \ldots, X_{i-1} \leq c_{i-1}, X_i > c_i, Y_i > d_i\right),$$

$$\mathcal{P}_O : \alpha_Y^O = \alpha_Y^S + \sum_{i=1}^{K_X \wedge \{K_Y - 1\}} \sum_{j=i+1}^{K_Y} \Pr\big(X_1 \leq c_1, \cdots, X_{i-1} \leq c_{i-1}, X_i > c_i,$$
$$Y_i \leq d_i, \cdots, Y_{j-1} \leq d_{j-1}, Y_j > d_j\big),$$

$$\mathcal{P}_P : \alpha_Y^P = \begin{cases} \alpha_Y^O + \Pr\big(X_1 \leq c_1, \cdots, X_{K_X} \leq c_{K_X}, Y_{K_Y} > d_{K_Y}\big), \text{ if} K_X \geq K_Y, \\ \alpha_Y^O + \sum_{i=K_X}^{K_Y} \Pr\big(X_1 \leq c_1, \cdots, X_{K_X} \leq c_{K_X}, Y_{K_X} \leq d_{K_X}, \ldots, \\ \qquad Y_{i-1} \leq d_{i-1}, Y_i > d_i\big), \text{ if} K_X < K_Y, \end{cases}$$

where $K_X \wedge K_Y = \min\{K_X, K_Y\}$.

Note that for a test on a primary and a secondary endpoint in a group sequential design, the control of familywise error rate (FWER) (Hochberg and Tamhane 1987; Tamhane et al. 2010; Zhang and Gou 2019a) requires that FWER $=$ $\Pr$(Reject at least one true $H \in \{H_X, H_Y\}$) $\leq \alpha$. Following the closure principle (Marcus et al. 1976), the control of type I error under primary hypothesis H_X, the control under secondary hypothesis H_Y and the control under their intersection $H_X \cap H_Y$ are all at level α, leading to the control of the FWER at level α.

11.3 Stagewise Hierarchical Rule

The stagewise hierarchical rule $\mathcal{P}_S$ and the overall hierarchical rule $\mathcal{P}_O$ satisfy the gatekeeping condition, In other words, the secondary endpoint is tested only if the primary endpoint is significant (Dmitrienko and Tamhane 2007; Dmitrienko et al. 2009). Under this condition, the event $R_Y = \{\text{Reject } H_Y\}$ is a subset of the event $R_X = \{\text{Reject } H_X\}$. It follows that $\Pr(R_X \cup R_Y | H_X \cap H_Y) = \Pr(R_X | H_X)$. This indicates that once the primary endpoint is tested using an α-level boundary, then $\Pr(R_X \cup R_Y | H_X \cap H_Y) \leq \alpha$ (Tamhane et al. 2010). Consequently, for testing procedures using the stagewise hierarchical rule $\mathcal{P}_S$ or the overall hierarchical rule $\mathcal{P}_O$, in order to control FWER at level α, the only requirement of type I error control for the secondary hypothesis is $\Pr(R_Y | H_Y) \leq \alpha$, or more specifically, $\Pr\big(R_Y | \overline{H}_X \cap H_Y\big) \leq \alpha$.

In a $[2|K]$-stage design, the primary hypothesis H_X can be tested flexibly using any α-level group sequential boundary (c_1, c_2). For example, the critical boundary (c_1, c_2) satisfies $\alpha_X = 1 - \Pr(X_1 \leq c_1, X_2 \leq c_2) \leq \alpha$. The marginal significance level of the secondary hypothesis H_Y is defined as $\alpha_Y = 1 - \Pr\big(\cap_{i=1}^K \{Y_i \leq d_i\}\big)$.

We consider using a more liberal secondary boundary $(d_1, \ldots, d_K)$ where α_Y can be greater than α with the control of FWER at level α.

Assume that the test statistics follow the multivariate normal distribution, which applies asymptotically to a wide range of test statistics. Namely,

$$\begin{pmatrix} X_1 \\ Y_1 \\ X_2 \\ Y_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \lambda\Delta_X \\ \gamma_1\Delta_Y \\ \Delta_X \\ \gamma_2\Delta_Y \end{pmatrix}, \begin{pmatrix} 1 & \frac{\min\{\lambda,\gamma_1\}}{\max\{\lambda,\gamma_1\}}\rho & \lambda & \frac{\min\{\lambda,\gamma_2\}}{\max\{\lambda,\gamma_2\}}\rho \\ \frac{\min\{\lambda,\gamma_1\}}{\max\{\lambda,\gamma_1\}}\rho & 1 & \gamma_1\rho & \gamma_1/\gamma_2 \\ \lambda & \gamma_1\rho & 1 & \gamma_2\rho \\ \frac{\min\{\lambda,\gamma_2\}}{\max\{\lambda,\gamma_2\}}\rho & \gamma_1/\gamma_2 & \gamma_2\rho & 1 \end{pmatrix}\right). \quad (11.1)$$

In the following, Theorem 1 gives an upper bound of type I error of stagewise hierarchical rule $\mathcal{P}_S$. Unlike the results where the primary and the secondary endpoint have the same information fractions (Tamhane et al. 2010; Glimm et al. 2010; Tamhane et al. 2018) or the results with only one interim analysis for the secondary hypothesis H_Y where $\gamma_2 = 1$ (Gou and Xi 2019), the upper bound we provided for multiple interim stages with different information fractions is not sharp. In other words, the following theorem guarantees a more liberal secondary boundary unconditionally.

Theorem 1 (Upper Bound for Type I Error) *When using a stagewise hierarchical rule $\mathcal{P}_S$ under H_Y, the type I error α_Y^S is bounded from above by*

$$\alpha_Y^S < 1 - \Pr(Y_1 \le d_1, Y_2 \le d_2).$$

When $0 < \gamma_1 < \gamma_2 < 1$, this upper bound cannot be achieved.

Specifically, when the primary hypothesis data are obtained earlier than the secondary hypothesis data, at stage 1 we have $n_{1,X} > n_{1,Y}$. It follows that the information fraction of the primary hypothesis at stage 1 is greater than the corresponding information fraction of the secondary hypothesis. Starting from Theorem 1 along with the assumption that $t_X > t_{1,Y}$ and the correlation ρ between X_2 and Y_K is positive, we show in Theorem 2 below that the type I error rate for a stagewise hierarchical test under the secondary hypothesis H_Y, or α_Y^S, is uniformly monotonous.

Theorem 2 (Uniform Monotonicity of Type I Error) *Consider two group sequential designs using $\mathcal{P}_S$, one with the square roots of information fractions $(\lambda, \gamma_1', \gamma_2)$ and boundaries (c_1, c_2, d_1', d_2'), and the other with $(\lambda, \gamma_1'', \gamma_2)$ and (c_1, c_2, d_1'', d_2''). Denote the corresponding type I errors under H_Y by $\alpha_Y^{S'}$ and $\alpha_Y^{S''}$, respectively. Suppose that these two designs share the same boundary for the primary hypothesis (c_1, c_2), and the same information fraction $t_X = \lambda^2$ at the interim analysis of the primary hypothesis and the information fraction $t_{2,Y} = \gamma_2^2$ at the second stage of the secondary hypothesis. If $\gamma_1' \le \gamma_1'' \le \lambda$, $d_1' \ge d_1''$ and*

$d_2' \geq d_2''$, *then for any* $\rho \in [0, 1]$ *and for any* Δ_X,

$$\alpha_Y^{S'} \leq \alpha_Y^{S''}.$$

In order to apply Theorem 2 to the OBF-POC design, where an OBF boundary is used for the primary endpoint and a POC boundary is used for the secondary endpoint, we need the following result. The OBF-POC design in the stagewise hierarchical rule is recommended by Tamhane et al. (2010, 2018) and Zhang and Gou (2019b).

Lemma 1 *Consider two trials that use the Pocock test with two stages under the same significance level. In one trial, the interim analysis is performed at information time* t', *and the corresponding Pocock boundary is* d'. *In the other trial, the interim analysis is performed at* t'' *with Pocock boundary* d''. *If* $t' < t''$, *then* $d' > d''$.

An immediate consequence of Theorem 2 and Lemma 1 is that, when the information fraction of the secondary hypothesis at the interim analysis is small compared to the information fraction of the primary hypothesis, the statistical power of group sequential design using the stagewise hierarchical rule will benefit greatly from the secondary boundary refinement. Formally, this means that the OBF-POC design with unrefined boundaries becomes more conservative for testing the secondary hypothesis H_Y when the information time at the first stage $t_{1,Y}$ becomes smaller.

Figure 11.1 shows that the error rate α_Y^S under H_Y of an OBF-POC design, where the α-level boundaries (c_1, c_2) and (d_1, d_2) are used, say, $\alpha = 1 - \Pr(X_1 \leq c_1, X_2 \leq c_2) = 1 - \Pr(Y_1 \leq d_1, Y_2 \leq d_2)$. Figure 11.1 confirms the result in Theorem 1 that the error rate α_Y^S is strictly less than α. It also confirms that the uniform monotonicity of α_Y^S as a function of $t_{1,Y}$ in Theorem 2. The error rate α_Y^S of an OBF-OBF design, where both primary and secondary boundary are OBF, is also bounded by α, and is uniformly monotonic of $t_{1,Y}$, as shown in Fig. 11.2. The boundary values (d_1, d_2) can be refined to allow α_Y^S to achieve α.

The secondary boundary can be refined without knowing the correlation ρ between two hypotheses by assuming the least favorable situation where $\rho = 1$. If ρ is known or can be estimated (Tamhane et al. 2012a,b), we can further refine the boundary for the secondary hypothesis. Table 11.1 gives an example of the refined boundary (d_1', d_2') of the secondary hypothesis using OBF-POC and OBF-OBF designs, where $\rho = 1, 0.8, 0.5$. The error rate α_Y^S equals the level of significance α exactly with the boundary refinement of the secondary hypothesis.

Since $\lim_{\Delta_X \to +\infty} \alpha_Y^S(\rho, \Delta_X) = \Pr(Y_1 > d_1)$, for any ρ, λ, γ_1 and γ_2, the refined secondary boundary d_1 in an OBF-POC design is at least z_α, where z_α is the upper α critical point of the standard normal distribution. Note that the naïve strategy in Hung et al. (2007), where the secondary boundary $d_1 = d_2 = z_\alpha$, has been shown to be liberal when the information fractions for the primary and the secondary endpoint are the same. Gou and Xi (2019) first observed that the naïve strategy in Hung et al. (2007) actually control the FWER when the primary and the secondary

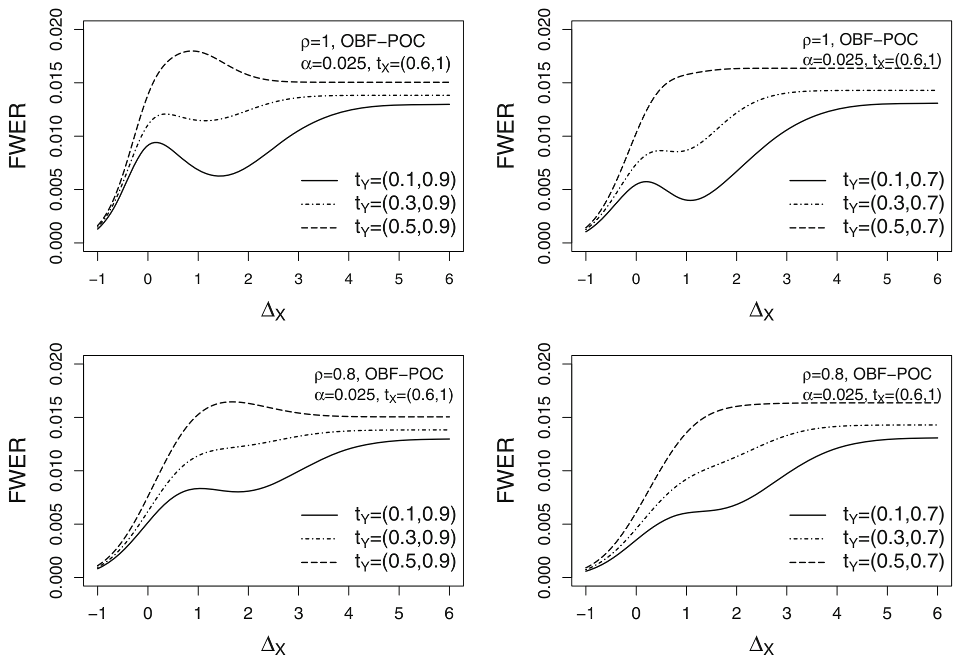

Fig. 11.1 FWER plot for O'Brien-Fleming primary and Pocock secondary boundary under $\mathcal{P}_S$ with $t_X = 0.6$, marginal level of significance $\alpha = 1 - \Pr(X_1 \leq c_1, X_2 \leq c_2) = 1 - \Pr(Y_1 \leq d_1, Y_2 \leq d_2) = 0.025$. Correlation $\rho = 1$ (top panels), $\rho = 0.8$ (bottom panels), $t_{2,Y} = 0.9$ (left panels), $t_{2,Y} = 0.7$ (right panels)

hypothesis have different information fractions, but without further discussion. A natural question here to ask is, when will the FWER inflation of the naïve strategy in Hung et al. (2007) not happen? Under an OBF-POC design, where an α-size OBF boundary (c_1, c_2) is chosen for the primary endpoint, and the boundary for the secondary endpoint is $d_1 = d_2 = z_\alpha$, Fig. 11.3 shows the admissible region of $(t_{1,Y}, t_{2,Y})$ for controlling the FWER of the naïve strategy in Hung et al. (2007) for different choices of the information fractions at the interim analysis of the primary hypothesis. The feasible region of $(t_{1,Y}, t_{2,Y})$ becomes larger when t_X increases. Generally speaking, when $(t_{1,Y}, t_{2,Y})$ are small enough compared with t_X, the naïve strategy controls the FWER. For example, in a phase III trial in Baselga et al. (2012), the primary endpoint is PFS with information fraction $\mathbf{t}_X = (0.6, 1)$, and the key secondary endpoint is OS with $\mathbf{t}_Y = (0.21, 0.44)$. If this trial follows the stagewise hierarchical strategy to control the FWER at level $\alpha = 0.025$ and uses an α-level OBF boundary for the PFS endpoint, then the boundary $d_1 = d_2 = z_\alpha = 1.960$ for the OS can be used since $t_{1,Y} = 0.21$ and $t_{2,Y} = 0.44$ fall into the admissible region when $t_X = 0.6$. This is shown in Fig. 11.3.

A simple empirical rule for properly using the naïve strategy in Hung et al. (2007) is followed: when $t_{1,Y}^2 \leq t_X$, a group sequential design with an 0.025-level OBF boundary for the primary hypothesis can directly apply $d_1 = d_2 = z_{0.025}$ as its boundary for the secondary hypothesis H_Y.

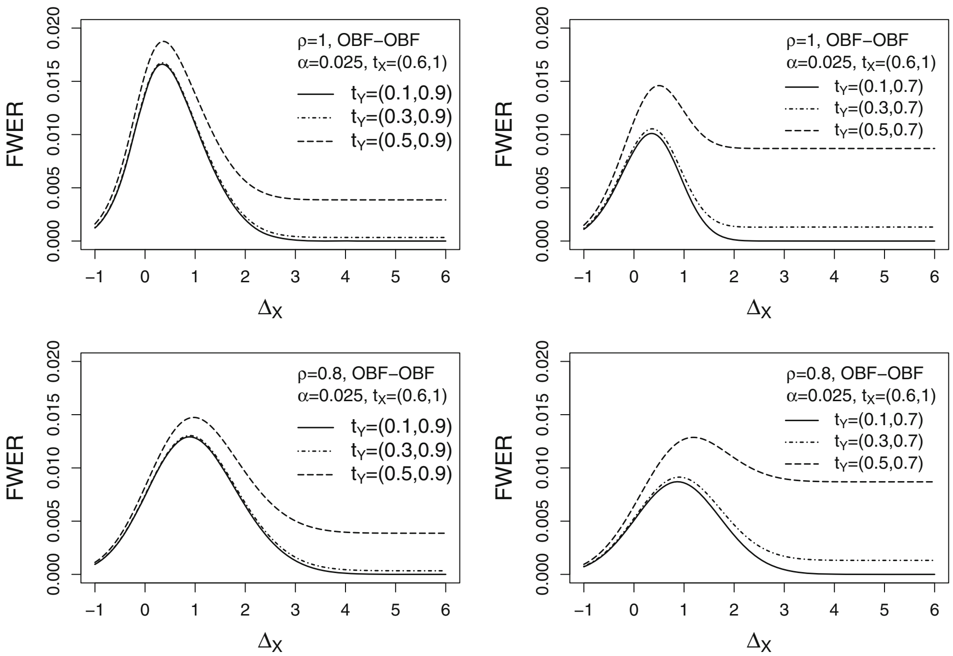

Fig. 11.2 FWER plot for O'Brien-Fleming primary and O'Brien-Fleming secondary boundary under $\mathcal{P}_S$ with $t_X = 0.6$, marginal level of significance $\alpha = 1 - \Pr(X_1 \le c_1, X_2 \le c_2) = 1 - \Pr(Y_1 \le d_1, Y_2 \le d_2) = 0.025$. Correlation $\rho = 1$ (top panels), $\rho = 0.8$ (bottom panels), $t_{2,Y} = 0.9$ (left panels), $t_{2,Y} = 0.7$ (right panels)

Table 11.1 Refined secondary boundaries for given correlation ρ under the stagewise hierarchical rule

OBF-POC	α-level boundary		Refined boundary		
ρ	d_1	d_2	d'_1	d'_2	Marginal error of H_Y
1	2.169	2.169	2.032	2.032	0.0345
0.8	2.169	2.169	1.996	1.996	0.0375
0.5	2.169	2.169	1.973	1.973	0.0394
OBF-OBF	**α-level boundary**		**Refined boundary**		
ρ	d_1	d_2	d'_1	d'_2	Marginal error of H_Y
1	2.664	1.985	2.511	1.872	0.0328
0.8	2.664	1.985	2.386	1.778	0.0408
0.5	2.664	1.985	2.308	1.721	0.0465

$t_X = 0.6$, $t_{1,Y} = 0.5$, $t_{2,Y} = 0.9$, the OBF boundary for the primary hypothesis is $c_1 = 2.572$, $c_2 = 1.992$ at $\alpha = 0.025$. The marginal error rate of H_Y is $1 - \Pr\left(Y_1 \le d'_1, Y_2 \le d'_2\right)$

In a $[K_X|K_Y]$-stage design following the stagewise hierarchical rule, similar conclusions on type I error rate can be achieved. The type I error rate α_Y^S is bounded from above by $1 - \Pr\left(Y_1 \le d_1, \ldots, Y_{K_X \wedge K_Y} \le d_{K_X \wedge K_Y}\right)$, and this upper bound is

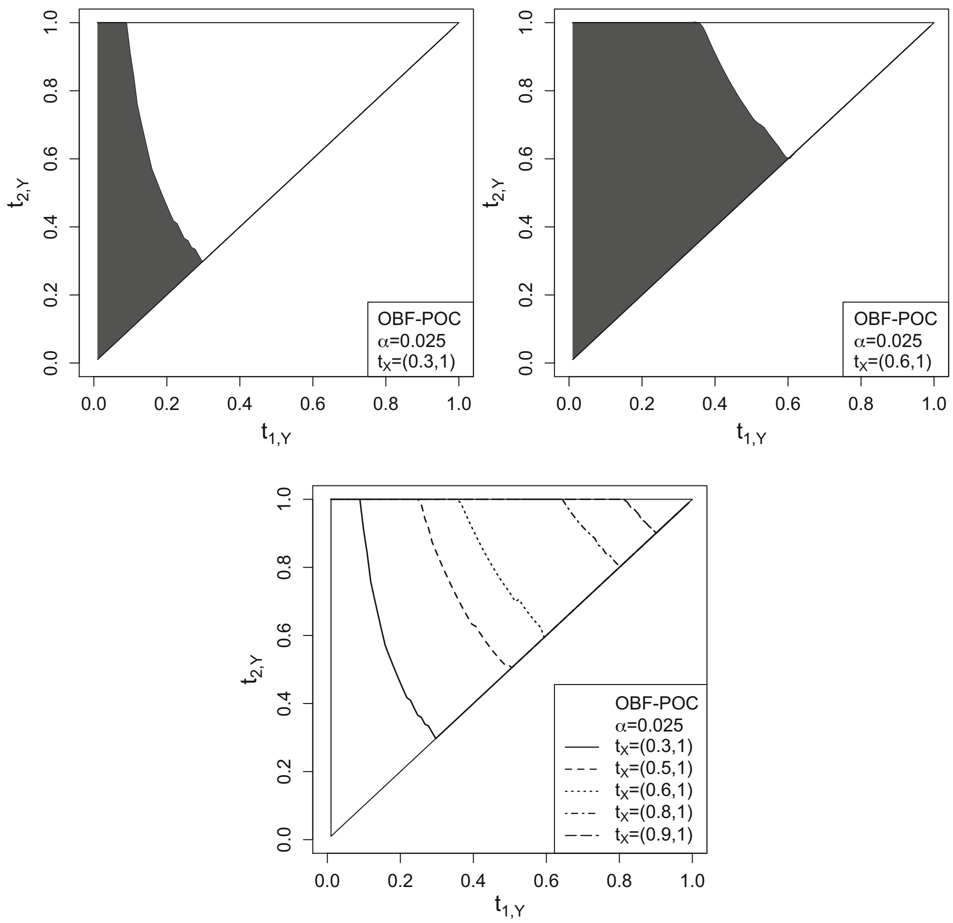

Fig. 11.3 Feasible region of $(t_{1,Y}, t_{2,Y})$ of the naïve strategy in Hung et al. (2007)

not sharp when $K_X \neq K_Y$. Under some conditions, the power gain for the secondary hypothesis H_Y by using the boundary refinement is significant when the information times of H_Y are less than the information times of the primary hypothesis H_X at interim stages.

11.4 Overall Hierarchical Rule

Compared with the stagewise hierarchical rule $\mathcal{P}_S$, a trial design using the overall hierarchical rule $\mathcal{P}_O$ allows testing the secondary hypothesis H_Y more than once if the primary hypothesis H_X is rejected. Following a similar argument in Tamhane et al. (2018), one cannot refine the secondary boundary unless there

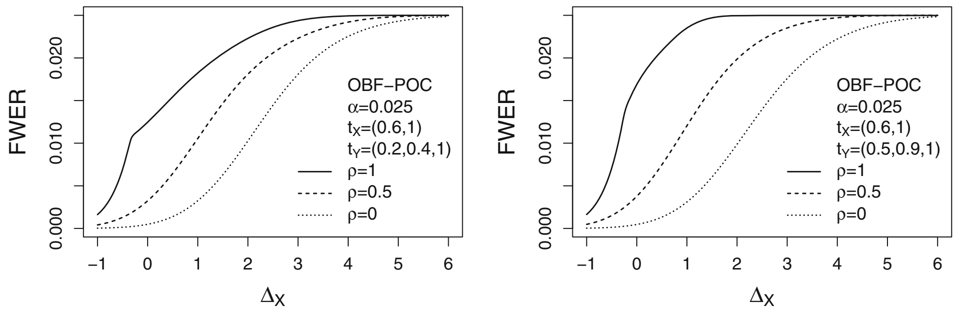

Fig. 11.4 FWER plot of an OBF-POC design using the overall hierarchical rule $\mathcal{P}_O$ with α-level boundary of the primary and the secondary hypothesis

is some prior information on Δ_X and ρ, since the difference between $1 - \Pr(Y_1 \le d_1, \cdots, Y_K \le d_K)$ and α_Y^O, which equals to

$$\Pr(X_1 \le c_1, X_2 \le c_2, Y_2 > d_2) + \Pr(X_1 \le c_1, Y_1 > c_1, Y_2 \le d_2, \cdots, Y_K \le d_K)$$
$$+ \sum_{i=2}^{K-1} \Pr(X_1 \le c_1, X_2 \le c_2, Y_2 \le d_2, \cdots, Y_i \le d_i, Y_{i+1} > d_{i+1}),$$

in a $[2|K]$-stage design, goes to 0 when Δ_X goes to positive infinity. Similarly, a $[K_X|K_Y]$-stage design using the overall hierarchical rule cannot be refined without information on Δ_X and ρ.

Figure 11.4 shows the type I error under H_Y of an OBF-POC design with $\alpha = 0.025$. Refinement of the secondary boundary is possible only when an upper bound on Δ_X is known. If a reliable estimate of Δ_X is available, the refinement of the boundary of the secondary hypothesis will be relatively noticeable when the time fraction of the secondary hypothesis $\mathbf{t}_Y$ is small or when the correlation ρ is small.

11.5 Partially Hierarchical Rule

The partially hierarchical rule $\mathcal{P}_P$ allows continued testing of the secondary hypothesis when the primary hypothesis has been confirmed to be non-significant. Thus, besides controlling of type I error under H_Y, one needs to also control the type I error under $H_X \cap H_Y$. Since $\Pr(R_X|H_X) \le \Pr(R_X \cup R_Y|H_X \cap H_Y)$, in general we cannot use an α-level significance for the primary endpoint in a design under the partially hierarchical rule.

A Bonferroni-based design splits the significance level α for H_X and H_Y whereby $\alpha = \alpha_X + \alpha_Y$, where $\alpha_X = 1 - \Pr(X_1 \le c_1, X_2 \le c_2)$ and $\alpha_Y = 1 - \Pr(Y_1 \le d_1, \cdots, Y_K \le d_K)$ in a $[2|K]$-stage design. This design controls the

FWER. When the correlation ρ is known or can be estimated, the boundary for the secondary hypothesis can be refined. The following theorem provides a refinement when ρ is known to be non-negative.

Theorem 3 (Improved Boundary in a [2|K]-Stage Design) *Consider a group sequential design using the partially hierarchical rule* $\mathcal{P}_P$*, where* $\alpha_X = \Pr(R_X|H_X) < \alpha$*. A necessary and sufficient condition for* $\Pr(R_X \cup R_Y|H_X \cap H_Y) \leq \alpha$ *for any non-negative* ρ *is that* $\Pr\left(\cap_{i=2}^{K}\{Y_i \leq d_i\}\right) \geq (1-\alpha)/(1-\alpha_X)$.

Based on Theorem 3, a simple design is followed when $\rho \geq 0$ is satisfied, where the boundary of the secondary hypothesis is refined.

1. $\alpha_X = 1 - \Pr(X_1 \leq c_1, X_2 \leq c_2) \leq \alpha$,
2. $\alpha_Y = 1 - \Pr(Y_1 \leq d_1, \cdots, Y_K \leq d_K) \leq \alpha$,
3. $1 - \Pr\left(\cap_{i=2}^{K}\{Y_i \leq d_i\}\right) \leq \frac{\alpha-\alpha_X}{1-\alpha_X}$.

Denote $\alpha_Y^{(-1)} = 1 - \Pr\left(\cap_{i=2}^{K}\{Y_i \leq d_i\}\right)$, which is the type I error of $(K-1)$-stage group sequential design for H_Y. Comparing the original K-stage group sequential design for H_Y, this $(K-1)$-stage design skips the first stage. Therefore, the conditions can be rewritten as

$$\alpha_X \leq \alpha, \quad \alpha_Y \leq \alpha, \quad \alpha_Y^{(-1)} \leq \frac{\alpha-\alpha_X}{1-\alpha_X}.$$

Theorem 3 can be easily generalized to a $[K_X|K_Y]$-stage design where the trial of the primary endpoint has more than two stages. The refined method maintains the FWER control across both endpoints.

Corollary 1 (Improved Boundary in a [$\mathbf{K_X}|K_Y$]-Stage Design) *Consider a group sequential design with* K_X *stages for the primary hypothesis and* K_Y *stages for the secondary hypothesis, using the partially hierarchical rule* $\mathcal{P}_P$*, where* $K_X \leq K_Y$*. Let* $\alpha_Y^{(-(K_X-1))} = 1 - \Pr\left(\cap_{i=K_X}^{K_Y}\{Y_i \leq d_i\}\right)$ *be the type I error of a* $(K_Y - K_X + 1)$*-stage group sequential design for* H_Y*. This procedure controls the FWER for arbitrary* $\rho \geq 0$ *if and only if:* $\alpha_X \leq \alpha$*,* $\alpha_Y \leq \alpha$*, and* $\alpha_Y^{(-(K_X-1))} \leq (\alpha-\alpha_X)/(1-\alpha_X)$.

The Lan-DeMets error spending function is widely used in clinical trials to approximate OBF and POC boundary (Lan and DeMets 1983). Using the Lan-DeMets boundaries, Table 11.2 shows the refined boundary for the secondary hypothesis under various values of ρ compared with the boundary based on Bonferroni split. The refined boundary for $\rho = 0$ can be used for any $\rho \geq 0$, based on Theorem 3. Even without the knowledge of the sign of ρ, the refined boundary for $\rho = 0$ is still approximately valid for endpoints with any correlation ρ, as shown in Table 11.2.

Table 11.2 Refined Lan-DeMets boundaries for the secondary hypothesis for given correlation ρ under the partially hierarchical rule $\mathcal{P}_P$, $\alpha = 0.025$, $\alpha_X = 0.0125$, $\mathbf{t}_X = (0.6, 1)$, $\mathbf{t}_Y = (0.5, 0.9, 1)$, Lan-DeMets OBF boundary for the primary hypothesis $(c_1, c_2) = (3.021, 2.254)$

	Lan-DeMets POC for the secondary				Lan-DeMets OBF for the secondary			
ρ	d_1	d_2	d_3	α_Y	d_1	d_2	d_3	α_Y
1	2.157	2.242	2.373	0.0250	2.963	2.105	2.057	0.0250
0.5	2.184	2.270	2.402	0.0234	3.237	2.313	2.249	0.0153
0	2.230	2.319	2.452	0.0208	3.304	2.363	2.295	0.0135
−0.5	2.235	2.324	2.457	0.0205	3.310	2.368	2.299	0.0133
Bonferroni	2.420	2.530	2.656	0.0125	3.345	2.394	2.323	0.0125

11.6 Power Analysis

In order to evaluate the performance of boundary refinement for the secondary hypothesis, we compare the secondary power $\Pr\left(R_Y|\overline{H}_Y\right)$ under the partially hierarchical rule $\mathcal{P}_P$ between the OBF-POC design and OBF-OBF design. Here, we only consider the O'Brien-Fleming boundary for the primary endpoint, since it is more powerful than the POC boundary for the primary hypothesis (Tamhane et al. 2018). For the power analysis, the assumption of multivariate normal distribution is satisfied asymptotically, so we incorporate the distribution information into the analysis. In general, if the distribution information is unknown, the power analysis models based on the Dirac function (Finner et al. 2009) or the step function (Zhang and Gou 2016) can be considered.

Table 11.3 displays the power comparisons between two designs (OBF-POC and OBF-OBF) and between two boundaries (refined boundary for $\rho \geq 0$ and unrefined boundary based on Bonferroni split). We assume the significance level $\alpha = 0.025$, and the primary hypothesis is tested with a 0.0125-level Lan-DeMets OBF boundary $(c_1, c_2) = (3.021, 2.254)$, where the information fraction at the interim analysis is 0.6. For the secondary hypothesis, we include the Lan-DeMets OBF and the Lan-DeMets POC boundary. Two choices of information fractions of the secondary endpoint show the impact of a fast data accumulation ($\mathbf{t}_Y = (0.5, 0.9, 1)$) and slow accumulation ($\mathbf{t}_Y = (0.2, 0.4, 1)$) for the secondary hypothesis. We assume the true correlation between the primary and the secondary hypothesis is 0.5. Note that we do not need to know this correlation for boundary refinement. The standardized treatment effect for the primary hypothesis Δ_X is 3, and it ranges from 2 to 4 for the secondary hypothesis.

From Table 11.3, we observe that the OBF-OBF design is better than the OBF-POC design in a group sequential trial using the partially hierarchical rule $\mathcal{P}_P$. Note that for a trial using the stagewise hierarchical rule $\mathcal{P}_S$, Tamhane et al. (2010), Tamhane et al. (2018) and Gou and Xi (2019) have shown that the OBF-POC is the better choice. For the OBF-OBF design using $\mathcal{P}_P$, the power gain over the Bonferroni split method increases when the information fractions of the secondary hypothesis become smaller.

Table 11.3 Power (%) comparison between the refined the unrefined boundary under the partially hierarchical rule

$\mathbf{t}_X = (0.6, 1)$		Lan-DeMets OBF-POC		Lan-DeMets OBF-OBF	
$\mathbf{t}_Y$	Δ_Y	Refined	Unrefined	Refined	Unrefined
(0.5, 0.9, 1)	2	39.1	31.7	40.7	39.5
	3	75.4	68.6	77.5	76.6
	4	95.2	92.8	96.0	95.7
(0.2, 0.4, 1)	2	38.5	34.1	44.8	40.4
	3	75.5	71.7	80.7	77.6
	4	95.4	94.1	96.9	96.1

11.7 Example and Extension

In practice, it is common that the attained sample sizes and the planned sample sizes are different. Using the error spending function, we can update the boundaries at each stage by considering the exact information fractions. The refined boundary can be updated in a similar manner adaptively.

Consider a phase III placebo-controlled two-arm clinical trial evaluating the efficacy of a treatment in patients with lymphoma. The primary objective is to evaluate the efficacy with respect to the progression-free survival (PFS). The secondary objective is to evaluate the efficacy with respect to the overall survival (OS). Table 11.4 shows a 0.025-level test using the partially hierarchical rule with a Lan-DeMets error spending function OBF-OBF design. The trial design includes one interim analysis for the primary endpoint PFS, and two interim analyses for the secondary endpoint OS. At stage 0, all sample sizes are planned. The sample size per arm is planned to be 400. The planned cumulative sample size for the primary objective is 240 at stage 1, and 400 at stage 2. For the secondary objective, the planned cumulative sample size is 200 at stage 1320 at stage 2, and 400 at stage 3. The critical boundaries for the primary and the secondary hypothesis can be calculated. At stage 1, $n_{1,X}$ and $n_{1,Y}$ are obtained, and the planned sample sizes for other stages are modified accordingly. The observed sample sizes at stage 1 for the primary and the secondary endpoint are 264 and 168, and the planned cumulative sample sizes at stage 2 and 3 remain the same. The critical boundary (c_1, c_2) and (d_1, d_2, d_3) are recalculated, and c_1 and d_1 are compared with the test statistics to make decisions. We further observe $n_{2,X}$ and $n_{2,Y}$ at stage 2, update the information times by using the observed cumulative sample sizes, and calculate the boundary c_2, c_3 and (d_2, d_3) by fixing the value of c_1 and c_2 in stage 1. Finally, $n_{3,Y}$ is observed at stage 3, and the total sample size for OS is updated, and the boundary d_3 is recalculated based on updated information times. In this example, initially the planned sample size is $(n_{1,X}, n_{2,X}, n_{1,Y}, n_{2,Y}, n_{3,Y}) = (240, 160, 200, 120, 80)$. At the final stage, the attained sample size becomes $(n_{1,X}, n_{2,X}, n_{1,Y}, n_{2,Y}, n_{3,Y}) = (264, 168, 168, 132, 108)$.

Table 11.4 Boundary updates among stages in an OBF-OBF design using $\mathcal{P}_P$: a comparison between the unrefined boundary (d_1, d_2, d_3) and the refined boundary (d'_1, d'_2, d'_3)

Stage	$n_{1,X}$	$n_{2,X}$	$n_{1,Y}$	$n_{2,Y}$	$n_{3,Y}$	c_1	c_2	d_1	d_2	d_3	d'_1	d'_2	d'_3
0	240	160	200	120	80	3.0205	2.2543	3.3446	2.5694	2.2938	3.2314	2.4794	2.2148
1	264	136	168	152	80	2.8614	2.2625	3.6810	2.5629	2.2928	3.5651	2.4770	2.2180
2	264	168	168	132	100	2.8614	2.2672	3.6810	2.6625	2.2835	3.5651	2.6529	2.2754
3	264	168	168	132	108	2.8614	2.2672	3.6810	2.6625	2.3170	3.5651	2.6529	2.3136

Acknowledgements We thank Ajit C. Tamhane and Dong Xi for comments that greatly improved the manuscript. This work was partially supported by the Professional Staff Congress-City University of New York (PSC-CUNY) research grant, Cycle 48 (2017–2018). This research article extended the framework that was present at the 2017 ICSA Applied Statistics Symposium, Session 148, Recent Developments in Theory and Application of Multiple Comparison Methods, *A gatekeeping test on a primary and a secondary endpoint in a group sequential design*, by Dr. Ajit C. Tamhane. It was also present at the 2017 ICSA Applied Statistics Symposium, Session 121, Multiplicity in Clinical Trials, *A gatekeeping test in a group sequential design with multiple interim looks*, by Dr. Jiangtao Gou. The authors thank editor Dr. Lanju Zhang and an anonymous referee for suggestions that improved this paper.

Conflict of Interest

The authors have declared no conflict of interest.

Appendix

Proof of Theorem 1 Note that $1 - \Pr(Y_1 \le d_1, Y_2 \le d_2) - \alpha_Y^S = \Pr(X_1 > c_1, Y_1 \le d_1, Y_2 > d_2) + \Pr(X_1 \le c_1, X_2 \le c_2, Y_2 > d_2) + \Pr(X_1 \le c_1, Y_1 > d_1, Y_2 \le d_2)$. All three terms on the right hand side are strictly positive when $\rho < 1$. When $\rho = 1$, the probability $\Pr(X_1 > c_1, Y_1 \le d_1, Y_2 > d_2)$ and $\Pr(X_1 \le c_1, Y_1 > d_1, Y_2 \le d_2)$ can be 0 if $\lambda = \gamma_1$, and the probability $\Pr(X_1 \le c_1, X_2 \le c_2, Y_2 > d_2)$ can be 0 if $\lambda = \gamma_2$. Since $\gamma_1 < \gamma_2$, these three terms cannot be 0 at the same time. It follows that $1 - \Pr(Y_1 \le d_1, Y_2 \le d_2)$ is strictly greater than α_Y^S. □

Proof of Theorem 2 Under $\overline{H}_X \cap H_Y$, the standardized treatment effects at the final stage for the secondary endpoint is zero, namely, $\Delta_Y = 0$. For simplicity, we denote the non-centrality parameters for the primary endpoint by $\Delta = \Delta_X$ under $\overline{H}_1 \cap H_2$. The type I error rate with smaller information fraction at stage 1 of the secondary hypothesis is

$$\begin{aligned}
\alpha_Y^{S\prime} &= \Pr\left(X_1 > c_1, Y_1' > d_1\right) + \Pr\left(X_1 \le c_1, X_2 > c_2, Y_2' > d_2\right) \\
&= \Pr\left(X_1 - \lambda\Delta > c_1 - \lambda\Delta, Y_1' > d_1\right) \\
&\quad + \Pr\left(X_1 - \lambda\Delta \le c_1 - \lambda\Delta, X_2 - \Delta > c_2 - \Delta, Y_2' > d_2\right)
\end{aligned}$$

For the first term, note that $\text{corr}\,(X_1, Y_1) = \gamma_1\rho/\lambda$. Since $\gamma_1' < \gamma_1''$ and $\rho \geq 0$, we have $\text{corr}\left(X_1, Y_1'\right) \leq \text{corr}\left(X_1, Y_1''\right)$. By Slepian's inequality (Plackett 1954; Slepian 1962) and $d_1' > d_1''$, it follows that

$$\Pr\left(X_1 - \lambda\Delta > c_1 - \lambda\Delta, Y_1' > d_1'\right) \leq \Pr\left(X_1 - \lambda\Delta > c_1 - \lambda\Delta, Y_1'' > d_1''\right).$$

For the second term, note that $\text{corr}\,(X_1, X_2) = \lambda$, $\text{corr}\,(X_2, Y_2) = \gamma_2\rho$, $\text{corr}\,(X_1, Y_2) = \rho \cdot \min\{\lambda, \gamma_2\}/\max\{\lambda, \gamma_2\}$, which are the same for the two designs. Since $d_2' \geq d_2''$, we get

$$\begin{aligned}&\Pr\left(X_1 - \lambda\Delta \leq c_1 - \lambda\Delta, X_2 - \Delta > c_2 - \Delta, Y_2 > d_2'\right)\\&\leq \Pr\left(X_1 - \lambda\Delta \leq c_1 - \lambda\Delta, X_2 - \Delta > c_2 - \Delta, Y_2 > d_2''\right).\end{aligned}$$

Thus $\alpha_Y^{S'} \leq \alpha_Y^{S''}$, for any $0 \leq \rho \leq 1$. □

Proof of Lemma 1 Suppose that $\left(Y_1', Y_2'\right)$ and $\left(Y_1'', Y_2''\right)$ are the bivariate normal distributed test statistics under the null hypothesis. The correlation between Y_1' and Y_2' is $\sqrt{t'}$, and the correlation between Y_1'' and Y_2'' is $\sqrt{t''}$. Since two trials have the same significance level α, we have

$$\Pr\left(Y_1' \leq d', Y_2' \leq d'\right) = 1 - \alpha = \Pr\left(Y_1'' \leq d'', Y_2'' \leq d''\right).$$

Since $\sqrt{t'} < \sqrt{t''}$, by Slepian's inequality, we get

$$\Pr\left(Y_1' \leq d', Y_2' \leq d'\right) < \Pr\left(Y_1'' \leq d', Y_2'' \leq d'\right).$$

It follows that

$$\Pr\left(Y_1'' \leq d', Y_2'' \leq d'\right) > \Pr\left(Y_1'' \leq d'', Y_2'' \leq d''\right).$$

Clearly, we have

$$d' > d''.$$

□

Proof of Theorem 3 For a design using the partially hierarchical rule $\mathcal{P}_P$, the error rate
$\Pr\,(R_X \cup R_Y | H_X \cap H_Y)$ is greater than $\Pr\,(R_X | H_X)$. The difference is bounded by

$$\begin{aligned}&\Pr\,(R_X \cup R_Y | H_X \cap H_Y) - \Pr\,(R_X | H_X)\\&= \Pr\,(X_1 \leq c_1, X_2 \leq c_2, Y_2 > d_2)\end{aligned}$$

$$+\sum_{i=2}^{K-1}\Pr(X_1 \le c_1, X_2 \le c_2, Y_2 \le d_2, \cdots, Y_i \le d_i, Y_{i+1} > d_{i+1})$$
$$= \Pr(X_1 \le c_1, X_2 \le c_2) - \Pr(X_1 \le c_1, X_2 \le c_2, Y_2 \le d_2, \cdots, Y_K \le d_K)$$
$$\le \Pr(X_1 \le c_1, X_2 \le c_2) - \Pr(X_1 \le c_1, X_2 \le c_2)\Pr\left(\cap_{i=2}^{K}\{Y_i \le d_i\}\right),$$

where $\Pr(X_1 \le c_1, X_2 \le c_2, Y_2 \le d_2, \cdots, Y_K \le d_K) \ge \Pr(X_1 \le c_1, X_2 \le c_2)$ $\Pr\left(\cap_{i=2}^{K}\{Y_i \le d_i\}\right)$ holds for any non-negative ρ, and two sides are equal when $\rho = 0$. It follows that

$$\Pr(R_X \cup R_Y | H_X \cap H_Y) - \alpha_X \le (1-\alpha_X)\left(1 - \Pr\left(\cap_{i=2}^{K}\{Y_i \le d_i\}\right)\right).$$

Also note that if

$$(1-\alpha_X)\left(1 - \Pr\left(\cap_{i=2}^{K}\{Y_i \le d_i\}\right)\right) \le \alpha - \alpha_X$$

the error rate control under intersection hypothesis, which is $\Pr(R_X \cup R_Y | H_X \cap H_Y) \le \alpha$, is guaranteed. Thus, if

$$\Pr\left(\cap_{i=2}^{K}\{Y_i \le d_i\}\right) \ge \frac{1-\alpha}{1-\alpha_X},$$

then $\Pr(R_X \cup R_Y | H_X \cap H_Y) \le \alpha$ for any $\rho \ge 0$. □

References

Amir, E., Seruga, B., Kwong, R., Tannock, I. F., Ocaña, A.: Poor correlation between progression-free and overall survival in modern clinical trials: are composite endpoints the answer? Eur. J. Cancer **48**, 385–388 (2012)

Baselga, J., Campone, M., Piccart, M., Burris III, H. A., Rugo, H. S., Sahmoud, T., Noguchi, S., Gnant, M., Pritchard, K. I., Lebrun, F., et al.: Everolimus in postmenopausal hormone-receptor–positive advanced breast cancer. N. Engl. J. Med. **366**, 520–529 (2012)

Bretz, F., Maurer, W., Brannath, W., Posch, M.: A graphical approach to sequentially rejective multiple test procedures. Stat. Med. **28**, 586–604 (2009)

Bretz, F., Posch, M., Glimm, E., Klinglmueller, F., Maurer, W., Rohmeyer, K.: Graphical approaches for multiple comparison procedures using weighted bonferroni, simes, or parametric tests. Biom. J. **53**, 894–913 (2011)

Burman, C.-F., Sonesson, C., Guilbaud, O.: A recycling framework for the construction of bonferroni-based multiple tests. Stat. Med. **28**, 739–761 (2009)

Dmitrienko, A., Tamhane, A.C.: Gatekeeping procedures with clinical trial applications. Pharm. Stat. **6**, 171–180 (2007)

Dmitrienko, A., Tamhane, A.C., Bretz, F.: Multiple Testing Problems in Pharmaceutical Statistics. Taylor & Francis, Boca Raton (2009)

Dunnett, C.W., Tamhane, A.C.: A step-up multiple test procedure. J. Am. Stat. Assoc. **87**, 162–170 (1992)
Finner, H., Dickhaus, T., Roters, M.: On the false discovery rate and an asymptotically optimal rejection curve. Ann. Stat. **37**, 596–618 (2009)
Fiteni, F., Westeel, V., Pivot, X., Borg, C., Vernerey, D., and Bonnetain, F.: Endpoints in cancer clinical trials. J. Visc. Surg. **151**, 17–22 (2014)
Glimm, E., Maurer, W., Bretz, F.: Hierarchical testing of multiple endpoints in group-sequential trials. Stat. Med. **29**, 219–228 (2010)
Gou, J., Tamhane, A.C. : On generalized Simes critical constants. Biom. J. **56**, 1035–1054 (2014)
Gou, J., Tamhane, A.C.: A flexible choice of critical constants for the improved hybrid Hochberg–Hommel procedure. Sankhya B **80**, 85–97 (2018a)
Gou, J., Tamhane, A.C.: Hochberg procedure under negative dependence. Stat. Sin. **28**, 339–362 (2018b)
Gou, J., Tamhane, A.C., Xi, D., Rom, D.: A class of improved hybrid Hochberg–Hommel type step-up multiple test procedures. Biometrika **101**, 899–911 (2014)
Gou, J., Xi, D.: Hierarchical testing of a primary and a secondary endpoint in a group sequential design with different information times. Stat. Biopharm. Res. (2019). https://doi.org/10.1080/19466315.2018.1546613
Hochberg, Y.: A sharper Bonferroni procedure for multiple tests of significance. Biometrika **75**, 800–802 (1988)
Hochberg, Y., Tamhane, A.C.: Multiple Comparison Procedures. Wiley, New York (1987)
Holm, S.: A simple sequentially rejective multiple test procedure. Scand. J. Stat. **6**, 65–70 (1979)
Hommel, G.: A stagewise rejective multiple test procedure based on a modified Bonferroni test. Biometrika **75**, 383–386 (1988)
Hung, H.M.J., Wang, S.-J., O'Neill, R.: Statistical considerations for testing multiple endpoints in group sequential or adaptive clinical trials. J. Biopharm. Stat. **17**, 1201–1210 (2007)
Jennison, C., Turnbull, B.W.: Group sequential tests for bivariate response: interim analyses of clinical trials with both efficacy and safety endpoints. Biometrics **49**, 741–752 (1993)
Jennison, C., Turnbull, B.W.: Group Sequential Methods with Applications to Clinical Trials. Chapman and Hall/CRC, New York (2000)
Lan, K.K.G., DeMets, D.L.: Discrete sequential boundaries for clinical trials. Biometrika **70**, 659–663 (1983)
Lan, K.K.G., DeMets, D.L.: Group sequential procedures: calendar versus information time. Stat. Med. **8**, 1191–1198 (1989)
Marcus, R., Peritz, E., Gabriel, K.R.: On closed testing procedures with special reference to ordered analysis of variance. Biometrika **63**, 655–660 (1976)
Maurer, W., Bretz, F.: Multiple testing in group sequential trials using graphical approaches. Stat. Biopharm. Res. **5**, 311–320 (2013)
Michiels, S., Saad, E.D., Buyse, M.: Progression-free survival as a surrogate for overall survival in clinical trials of targeted therapy in advanced solid tumors. Drugs **77**, 713–719 (2017)
O'Brien, P.C., Fleming, T.R.: A multiple testing procedure for clinical trials. Biometrics **35**, 549–556 (1979)
Plackett, R.L.: A reduction formula for normal multivariate integrals. Biometrika **41**, 351–360 (1954)
Pocock, S.J.: Group sequential methods in the design and analysis of clinical trials. Biometrika **64**, 191–199 (1977)
Rom, D.M.: A sequentially rejective test procedure based on a modified Bonferroni inequality. Biometrika **77**, 663–665 (1990)
Sarkar, S.K.: Generalizing Simes' test and Hochberg's stepup procedure. Ann. Stat. **36**, 337–363 (2008)
Simes, R.J.: An improved Bonferroni procedure for multiple tests of significance. Biometrika **73**, 751–754 (1986)
Slepian, D.: The one-sided barrier problem for gaussian noise. Bell Syst. Tech. J. **41**, 463–501 (1962)

Tamhane, A.C., Gou, J.: Advances in p-value based multiple test procedures. J. Biopharm. Stat. **28**, 10–27 (2018)

Tamhane, A.C., Gou, J., Jennison, C., Mehta, C.R., Curto, T.: A gatekeeping procedure to test a primary and a secondary endpoint in a group sequential design with multiple interim looks. Biometrics **74**, 40–48 (2018)

Tamhane, A.C., Mehta, C.R., Liu, L.: Testing a primary and a secondary endpoint in a group sequential design. Biometrics **66**, 1174–1184 (2010)

Tamhane, A.C., Wu, Y., Mehta, C.R.: Adaptive extensions of a two-stage group sequential procedure for testing primary and secondary endpoints (I): unknown correlation between the endpoints. Stat. Med. **31**, 2027–2040 (2012a)

Tamhane, A.C., Wu, Y., Mehta, C.R.: Adaptive extensions of a two-stage group sequential procedure for testing primary and secondary endpoints (II): sample size re-estimation. Stat. Med. **31**, 2041–2054 (2012b)

Tang, D.-I., Geller, N.L.: Closed testing procedures for group sequential clinical trials with multiple endpoints. Biometrics **55**, 1188–1192 (1999)

Xi, D., Tamhane, A.C.: Allocating recycled significance levels in group sequential procedures for multiple endpoints. Biom. J. **57**, 90–107 (2015)

Ye, Y., Li, A., Liu, L., Yao, B.: A group sequential holm procedure with multiple primary endpoints. Stat. Med. **32**, 1112–1124 (2013)

Zhang, F., Gou, J.: A p-value model for theoretical power analysis and its applications in multiple testing procedures. BMC Med. Res. Methodol. **16**, 135 (2016)

Zhang, F., Gou, J.: Control of false positive rates in clusterwise fMRI inferences. J. Appl. Stat. (2019a). https://doi.org/10.1080/02664763.2019.1573883

Zhang, F., Gou, J.: Refined critical boundary with enhanced statistical power for non-directional two-sided tests in group sequential designs with multiple endpoints. (2019b) (submitted)