

# Adaptive extensions of a two-stage group sequential procedure for testing primary and secondary endpoints (II): sample size re-estimation

Ajit C. Tamhane,<sup>a,\*†</sup> Yi Wu<sup>b</sup> and Cyrus R. Mehta<sup>c,d</sup>

In this part II of the paper on adaptive extensions of a two-stage group sequential procedure (GSP) for testing primary and secondary endpoints, we focus on the second stage sample size re-estimation based on the first stage data. First, we show that if we use the Cui–Huang–Wang statistics at the second stage, then we can use the same primary and secondary boundaries as for the original procedure (without sample size re-estimation) and still control the type I familywise error rate. This extends their result for the single endpoint case. We further show that the secondary boundary can be sharpened in this case by taking the unknown correlation coefficient  $\rho$  between the primary and secondary endpoints into account through the use of the confidence limit method proposed in part I of this paper. If we use the sufficient statistics instead of the CHW statistics, then we need to modify both the primary and secondary boundaries; otherwise, the error rate can get inflated. We show how to modify the boundaries of the original group sequential procedure to control the familywise error rate. We provide power comparisons between competing procedures. We illustrate the procedures with a clinical trial example. Copyright © 2012 John Wiley & Sons, Ltd.

**Keywords:** adaptive designs; familywise error rate; gatekeeping procedures; multiple comparisons; multiple endpoints; O’Brien–Fleming boundary; Pocock boundary; sample size re-estimation.

## 1. Introduction

Adaptive designs have received wide acceptance both from clinical researchers as well as regulators. A common adaptation is sample size re-estimation based on interim estimates of the treatment effect and/or of the variance of the observations. Much work has been carried out on the problem of sample size re-estimation in group sequential designs; see, for example, [1–7]. All of these works deal with a single endpoint. In this paper, we extend these methods to the two endpoints case where the primary endpoint is a gatekeeper for the secondary endpoint and the two-stage GSP proposed in [8] is used.

The outline of the paper is as follows. In Section 2, we review the basic notation and assumptions. Section 3 describes two methods of sample size re-estimation. Section 4 studies two GSPs based on the Cui–Hung–Wang (CHW) statistics. The first GSP uses separate  $\alpha$ -level primary and secondary boundaries; this choice of boundaries implicitly assumes the least favorable value of  $\rho = 1$ . The second GSP uses sharper boundaries that take into account the unknown  $\rho$  through the confidence limit method proposed in part I. We provide a power comparison between the two GSPs to assess the power advantage of the second GSP. Section 5 discusses the use of the sufficient test statistics instead of the CHW test statistics. Section 6 discusses numerical evaluation of the critical boundaries for the GSP based on the sufficient test statistics. Section 7 gives simulation-based power comparisons between competing

<sup>a</sup>Department of IEMS, Northwestern University, Evanston, IL 60208, U.S.A.

<sup>b</sup>Department of Statistics, Northwestern University, Evanston, IL 60208, U.S.A.

<sup>c</sup>Cytel Corporation, Cambridge, MA 02139, U.S.A.

<sup>d</sup>Department of Biostatistics, Harvard School of Public Health, Cambridge, MA, U.S.A.

\*Correspondence to: Ajit C. Tamhane, Department of IEMS, Northwestern University, Evanston, IL 60208, U.S.A.

†E-mail: atamhane@northwestern.edu

procedures. Section 8 shows how the methods presented in the previous sections for the one-sample problem extend to the two-sample problem. Section 9 gives a clinical trial example to illustrate the methodology proposed in the paper. Finally, Section 10 gives a discussion and final recommendation.

## 2. Notation and background

The basic setup is the same as in part I [9], which we review here briefly for convenience. Consider a two-stage GSP with sample sizes  $n_1$  and  $n_2$  with independent and identically distributed bivariate normal observations  $(X_{ij}, Y_{ij})$  on the primary and secondary endpoints for the  $j$ th patient in the  $i$ th stage ( $i = 1, 2, j = 1, \dots, n_i$ ), where  $X_{ij} \sim N(\mu_1, \sigma_1^2)$ ,  $Y_{ij} \sim N(\mu_2, \sigma_2^2)$  and  $\text{corr}(X_{ij}, Y_{ij}) = \rho \geq 0$ . Let  $\delta_1 = \mu_1/\sigma_1$  and  $\delta_2 = \mu_2/\sigma_2$ . For convenience, assume that  $\sigma_1$  and  $\sigma_2$  are known, but all other parameters are unknown. In practice,  $\sigma_1$  and  $\sigma_2$  are unknown and are estimated. In Section 5 of part I, we studied via simulation the effect on the familywise error rate (FWER) of using estimates in place of the unknown  $\sigma_1$  and  $\sigma_2$  and found that the FWER requirement (1) is satisfied.

The hypotheses to be tested are  $H_1 : \delta_1 = 0$  and  $H_2 : \delta_2 = 0$  against upper one-sided alternatives, subject to the gatekeeping restriction that  $H_2$  can be tested only if  $H_1$  is rejected; otherwise,  $H_2$  is accepted without a test. Any test procedure under consideration must satisfy the FWER requirement

$$\text{FWER} = P\{\text{Reject at least one true } H_i (i = 1, 2)\} \leq \alpha \quad (1)$$

for a specified  $\alpha$  when either  $H_1$  or  $H_2$  is true.

The first stage test statistics are defined as

$$X_1 = X^{(1)} = \frac{\bar{X}^{(1)}}{\sigma_1/\sqrt{n_1}} \text{ and } Y_1 = Y^{(1)} = \frac{\bar{Y}^{(1)}}{\sigma_2/\sqrt{n_1}}, \quad (2)$$

where  $\bar{X}^{(1)}$  and  $\bar{Y}^{(1)}$  are the first stage sample means. Similarly, the second stage test statistics are defined as

$$X^{(2)} = \frac{\bar{X}^{(2)}}{\sigma_1/\sqrt{n_2}} \text{ and } Y^{(2)} = \frac{\bar{Y}^{(2)}}{\sigma_2/\sqrt{n_2}}, \quad (3)$$

where  $\bar{X}^{(2)}$  and  $\bar{Y}^{(2)}$  are the second stage sample means. Then, the cumulative test statistics at the final look can be expressed as

$$X_2 = \sqrt{f}X^{(1)} + \sqrt{1-f}X^{(2)}, Y_2 = \sqrt{f}Y^{(1)} + \sqrt{1-f}Y^{(2)} \quad (4)$$

where  $f = n_1/(n_1 + n_2)$  is the information fraction [10] at the interim look.

The GSP, denoted by  $\mathcal{P}$ , operates as follows.

Stage 1. Take  $n_1$  observations,  $(X_{1j}, Y_{1j})$ ,  $j = 1, \dots, n_1$ , and compute  $(X_1, Y_1)$ . If  $X_1 \leq c_1$ , continue to stage 2. If  $X_1 > c_1$ , reject  $H_1$  and test  $H_2$ . If  $Y_1 > d_1$ , reject  $H_2$ ; otherwise, accept  $H_2$ . In either case, stop sampling.

Stage 2. Take  $n_2$  observations,  $(X_{2j}, Y_{2j})$ ,  $j = 1, \dots, n_2$ , and compute  $(X_2, Y_2)$ . If  $X_2 \leq c_2$ , accept  $H_1$  and stop testing; otherwise, reject  $H_1$  and test  $H_2$ . If  $Y_2 > d_2$ , reject  $H_2$ ; otherwise, accept  $H_2$ .

We determine the critical boundaries  $(c_1, c_2)$  and  $(d_1, d_2)$  of  $\mathcal{P}$  to satisfy the FWER requirement (1).

We next define the notation for the case where the second stage sample size is re-estimated in light of the first stage data. We will denote the re-estimated second stage sample size by  $n'_2$ ; typically,  $n'_2 \geq n_2$ . Rules for determining  $n'_2$  are generally functions of the primary conditional power (CP) of the GSP, defined as

$$\text{CP} = P\{X_2 > c_2 | X_1 = x_1\}. \quad (5)$$

In the sequel, we will consider two different statistics for adaptive designs. To define these statistics, denote the second stage test statistics for the primary and secondary endpoints based on  $n'_2$  by

$$X'^{(2)} = \frac{\bar{X}'^{(2)}}{\sigma_1/\sqrt{n'_2}} \text{ and } Y'^{(2)} = \frac{\bar{Y}'^{(2)}}{\sigma_2/\sqrt{n'_2}}, \quad (6)$$

where  $\bar{X}'^{(2)}$  and  $\bar{Y}'^{(2)}$  are the second stage sample means based on  $n'_2$ . The CHW statistics are defined as

$$X'_2 = \sqrt{f}X^{(1)} + \sqrt{1-f}X'^{(2)} \text{ and } Y'_2 = \sqrt{f}Y^{(1)} + \sqrt{1-f}Y'^{(2)}, \quad (7)$$

which use the unadjusted weights  $\sqrt{f}$  and  $\sqrt{1-f}$  based on the preplanned sample sizes. As a result, they are not sufficient statistics, that is, they are not functions of the overall sample means  $\bar{X}'_2 = (n_1\bar{X}^{(1)} + n'_2\bar{X}'^{(2)}) / (n_1 + n'_2)$  and  $\bar{Y}'_2 = (n_1\bar{Y}^{(1)} + n'_2\bar{Y}'^{(2)}) / (n_1 + n'_2)$ , respectively. Therefore, we will also consider the following sufficient test statistics at the second stage:

$$X''_2 = \frac{\bar{X}'_2}{\sigma_1/\sqrt{n_1 + n'_2}} = \sqrt{f'}X^{(1)} + \sqrt{1-f'}X'^{(2)} \quad (8)$$

and

$$Y''_2 = \frac{\bar{Y}'_2}{\sigma_2/\sqrt{n_1 + n'_2}} = \sqrt{f'}Y^{(1)} + \sqrt{1-f'}Y'^{(2)}, \quad (9)$$

where  $f' = n_1/(n_1 + n'_2)$  is the information fraction at the interim look for the adaptive design and  $\sqrt{f'}$  and  $\sqrt{1-f'}$  are the adjusted weights based on the adjusted sample size.

### 3. Methods for sample size re-estimation

The second stage sample size will be re-estimated if the primary CP after the first stage is not large enough to meet the prespecified power requirement. Generally, this is because the actual treatment effect is smaller than the expected treatment effect for which the trial was designed to meet the prespecified power requirement. An expression for CP is given by

$$\begin{aligned} \text{CP} &= P\left(\sqrt{f}X^{(1)} + \sqrt{1-f}X^{(2)} > c_2 \mid X^{(1)} = x_1\right) \\ &= P\left(X^{(2)} - \delta_1\sqrt{n_2} > \frac{c_2 - x_1\sqrt{f}}{\sqrt{1-f}} - \delta_1\sqrt{n_2} \mid X^{(1)} = x_1\right) \\ &= 1 - \Phi\left(\frac{c_2 - x_1\sqrt{f}}{\sqrt{1-f}} - \delta_1\sqrt{n_2}\right). \end{aligned} \quad (10)$$

We will consider two different rules for determining the second stage adjusted sample size  $n'_2$ .

1. Gao *et al.* [6] proposed determining  $n'_2$  so that CP is preserved at the planned value  $1 - \beta$ . If  $z_\beta$  denotes the  $1 - \beta$  quantile of the standard normal distribution, then from (10), we get the following formula for  $n'_2$ :

$$n'_2 = \frac{1}{\delta_1^2} \left\{ \frac{1}{\sqrt{n_2}} [c_2\sqrt{n_1 + n_2} - x_1\sqrt{n_1}] + z_\beta \right\}^2. \quad (11)$$

In this expression,  $\delta_1$  is an unknown parameter, which we can replace by its sample estimate,  $\hat{\delta}_1 = x_1/\sqrt{n_1}$ . However, this can result in impractically large  $n'_2$  values if  $x_1$  or equivalently CP is small. Therefore, we truncate  $n'_2$  using the following modification:

$$n'_2 = \begin{cases} \min(n'_2 \text{ according to (11), } \gamma n_2) & \text{if } \text{CP}_{\min} \leq \text{CP} \leq \text{CP}_{\max} \\ n_2 & \text{otherwise,} \end{cases} \quad (12)$$

where  $[\text{CP}_{\min}, \text{CP}_{\max}]$  is called the *promising zone* and  $\gamma > 1$  is a prespecified constant.

2. Because of the sensitivity of the aforementioned formula to  $\hat{\delta}_1$  and the numerical difficulties involved in computing the modified critical boundaries, we carry out the power comparisons reported in the present paper by using a simpler version of the aforementioned rule, called the *fixed increase* rule in which we drop the dependence of  $n'_2$  on  $\hat{\delta}_1$ . Thus,

$$n'_2 = \begin{cases} \gamma n_2 & \text{if } \text{CP}_{\min} \leq \text{CP} \leq \text{CP}_{\max} \\ n_2 & \text{otherwise.} \end{cases} \quad (13)$$

It is possible to devise other modifications. For example, one could increase  $n_2$  to  $n'_2 = \gamma n_2$  for CP in the interval  $[CP_{\min}, CP_{\text{mid}}]$ , where  $CP_{\text{mid}}$  is some point in the interval  $[CP_{\min}, CP_{\max}]$  (e.g., the midpoint of the interval) and linearly taper off  $n'_2$  to  $n_2$  over the interval  $[CP_{\text{mid}}, CP_{\max}]$ .

#### 4. Procedures based on Cui–Hung–Wang statistics

We will offer two procedures, labeled  $\mathcal{P}'_1$  and  $\mathcal{P}'_2$ , based on the CHW statistics.  $\mathcal{P}'_1$  uses separate  $\alpha$ -level boundaries,  $(c_1, c_2)$  and  $(d_1, d_2)$ , for testing  $H_1$  and  $H_2$  (see (14)), and implicitly assumes that the unknown  $\rho$  is at its least favorable value  $\rho = 1$ . On the other hand,  $\mathcal{P}'_2$  uses the adjusted secondary boundary  $(d'_1, d'_2)$  on the basis of the confidence limit method for unknown  $\rho$  proposed in part I.

##### 4.1. Procedure $\mathcal{P}'_1$

Let  $(c_1, c_2)$  and  $(d_1, d_2)$  be any  $\alpha$ -level boundaries for the primary and secondary endpoints, respectively. Procedure  $\mathcal{P}'_1$  is the same as  $\mathcal{P}$  stated in Section 2 but uses the CHW statistics  $(X_1, X'_2)$  and  $(Y_1, Y'_2)$  in conjunction with these boundaries. In Proposition 1, we show that  $\mathcal{P}'_1$  controls the FWER. In fact, we will show this result for a general  $m$ -stage GSP for  $m \geq 2$ , which is an extension of the corresponding result in CHW for the single endpoint case. We first introduce the necessary notation for this general case.

Consider a non-adaptive procedure  $\mathcal{P}$  for a general  $m$ -stage GSP with sample sizes  $n_1, \dots, n_m$ , which uses the test statistics  $(X_1, \dots, X_m)$  and  $(Y_1, \dots, Y_m)$  in conjunction with the boundaries  $(c_1, \dots, c_m)$  and  $(d_1, \dots, d_m)$ , respectively. The test statistics  $(X_i, Y_i)$  are the standardized cumulative sample means at the  $i$ th look ( $i = 1, \dots, m$ ) and  $(c_1, \dots, c_m)$  and  $(d_1, \dots, d_m)$  are separate  $\alpha$ -level boundaries, which satisfy

$$P_{H_1}(\cup_{i=1}^m \{X_i > c_i\}) = \alpha \text{ and } P_{H_2}(\cup_{i=1}^m \{Y_i > d_i\}) = \alpha. \tag{14}$$

Consider a similar generalization of the adaptive  $m$ -stage GSP in which the sample sizes are adjusted in the same proportion at some stage  $\ell < m$  for all subsequent stages  $i > \ell$ , that is, the adjusted sample sizes are  $n'_i = \gamma n_i$  for all stages  $i > \ell$ , where the common proportionality constant  $\gamma \geq 1$  is a function of the data up to stage  $\ell$ . Let  $N_i = n_1 + \dots + n_i$  be the cumulative unadjusted sample sizes for stages  $i = 1, \dots, m$  and let  $f_{ij} = n_j/N_i$  ( $1 \leq j \leq i, 1 \leq i \leq m$ ). Denote by  $X^{(i)}$  and  $Y^{(i)}$  the stagewise standardized statistics for stage  $i$  in the non-adaptive part of the procedure where

$$X^{(i)} = \frac{\sum_{j=1}^{n_i} X_{ij}}{\sigma_1 \sqrt{n_i}} \text{ and } Y^{(i)} = \frac{\sum_{j=1}^{n_i} Y_{ij}}{\sigma_2 \sqrt{n_i}} \text{ for } 1 \leq i \leq \ell$$

and by  $X'^{(i)}$  and  $Y'^{(i)}$  the corresponding statistics for the adaptive part of the procedure, where

$$X'^{(i)} = \frac{\sum_{j=1}^{n'_i} X_{ij}}{\sigma_1 \sqrt{n'_i}} \text{ and } Y'^{(i)} = \frac{\sum_{j=1}^{n'_i} Y_{ij}}{\sigma_2 \sqrt{n'_i}} \text{ for } \ell + 1 \leq i \leq m.$$

Then, the CHW statistics for stages  $i > \ell$  can be expressed as

$$X'_i = \sum_{j=1}^{\ell} \sqrt{f_{ij}} X^{(j)} + \sum_{j=\ell+1}^i \sqrt{f_{ij}} X'^{(j)} \text{ and } Y'_i = \sum_{j=1}^{\ell} \sqrt{f_{ij}} Y^{(j)} + \sum_{j=\ell+1}^i \sqrt{f_{ij}} Y'^{(j)}, \tag{15}$$

which generalize those in (7). These CHW statistics can be shown to be equivalent to Fisher's [1] variance function statistics when the sample size re-estimation is carried out only at one stage during the trial and the sample sizes for all subsequent stages are adjusted in the same proportion.

##### Proposition 1

Consider an  $m$ -stage group sequential procedure  $\mathcal{P}$  with fixed sample sizes  $n_1, \dots, n_m$ , which uses the test statistics  $(X_1, \dots, X_m)$  and  $(Y_1, \dots, Y_m)$  in conjunction with  $\alpha$ -level boundaries  $(c_1, \dots, c_m)$  and  $(d_1, \dots, d_m)$  satisfying (14). Let  $\mathcal{P}'_1$  be an adaptive extension of  $\mathcal{P}$ , which adjusts all sample sizes after stage  $\ell < m$  in the same proportion so that  $n'_i = \gamma n_i$  for some  $\gamma \geq 1$  for all  $i > \ell$  and uses the statistics  $(X_1, \dots, X_\ell, X'_{\ell+1}, \dots, X'_m)$  and  $(Y_1, \dots, Y_\ell, Y'_{\ell+1}, \dots, Y'_m)$  in conjunction with the same  $\alpha$ -level

boundaries  $(c_1, \dots, c_m)$  and  $(d_1, \dots, d_m)$ , where  $(X'_i, Y'_i)$  for  $\ell + 1 \leq i \leq m$  are the CHW statistics defined in (15). Then,  $\mathcal{P}'_1$  controls the FWER at level  $\alpha$ .

*Proof*

We will first show that  $\mathcal{P}$  controls the FWER at level  $\alpha$ . Consider three cases in which a type I error can occur.

*Case 1 (Both  $H_1$  and  $H_2$  are true):* In this case,

$$\begin{aligned} \text{FWER} &= P_{H_1 \cap H_2}(\{\text{Reject } H_1\} \cup \{\text{Reject } H_2\}) \\ &= P_{H_1}(\text{Reject } H_1) \text{ (because } H_2 \text{ can be rejected only if } H_1 \text{ is rejected)} \\ &= P_{H_1}(\cup_{i=1}^m \{X_i > c_i\}) = \alpha. \end{aligned}$$

*Case 2 ( $H_1$  is true and  $H_2$  is false):* In this case,

$$\text{FWER} = P_{H_1}(\text{Reject } H_1) = P_{H_1}(\cup_{i=1}^m \{X_i > c_i\}) = \alpha.$$

*Case 3 ( $H_1$  is false and  $H_2$  is true):* In this case,

$$\begin{aligned} \text{FWER} &= P_{H_2}(\text{Reject } H_2) \\ &= P_{H_2}(\cup_{i=1}^m \{X_1 \leq c_1, \dots, X_{i-1} \leq c_{i-1}, X_i > c_i, Y_i > d_i\}) \\ &< P_{H_2}(\cup_{i=1}^m \{Y_i > d_i\}) = \alpha. \end{aligned}$$

The same proofs go through for  $\mathcal{P}'_1$  because, as shown by Cui *et al.* [2], the joint distributions of  $(X_1, \dots, X_\ell, X'_{\ell+1}, \dots, X'_m)$  and  $(X_1, \dots, X_\ell, X_{\ell+1}, \dots, X_m)$  are the same under  $H_1$ . Similarly, the joint distributions of  $(Y_1, \dots, Y_\ell, Y'_{\ell+1}, \dots, Y'_m)$  and  $(Y_1, \dots, Y_\ell, Y_{\ell+1}, \dots, Y_m)$  are the same under  $H_2$ .  $\square$

*Remark 1*

The aforementioned result for  $m = 2$  follows directly from Propositions 1–3 in [8], but it was not explicitly stated. The proof given there was more complicated. The present proof due to Liu [11] is more direct and simpler besides being more general.

4.2. Procedure  $\mathcal{P}'_2$

As can be seen from the aforementioned proof,  $\mathcal{P}'_1$  does not use any information on  $\rho$ ; effectively, it assumes the least favorable value  $\rho = 1$ . We now propose procedure  $\mathcal{P}'_2$  that uses the confidence limit method of part I based on the first stage sample correlation coefficient  $r$  to obtain a sharper second stage boundary.

*Proposition 2*

Let  $(c_1, c_2)$  and  $(d_1, d_2)$  be any  $\alpha$ -level boundaries of  $\mathcal{P}$  that control the FWER at level  $\alpha$ . Then,  $\mathcal{P}'_2$ , which uses the CHW statistics in conjunction with the critical boundaries  $(c'_1, c'_2)$  and  $(d'_1, d'_2)$ , controls the FWER at level  $\alpha$  if we set  $(c'_1, c'_2) = (c_1, c_2)$ ,  $d'_1 = d_1$  and  $d'_2$  as the solution to the equation

$$\Phi_2 \left( \frac{\sqrt{f}x_1 - c_2}{\sqrt{1-f}} + \delta_1 \sqrt{n'_2}, \frac{\sqrt{f}y_1 - d'_2}{\sqrt{1-f}} \middle| \rho \right) = \Phi_2 \left( \frac{\sqrt{f}x_1 - c_2}{\sqrt{1-f}} + \delta_1 \sqrt{n_2}, \frac{\sqrt{f}y_1 - d_2}{\sqrt{1-f}} \middle| \rho \right), \quad (16)$$

where  $(x_1, y_1)$  are the observed values of  $(X_1, Y_1)$  and  $\Phi_2(\cdot, \cdot | \rho)$  is the CDF of the standard bivariate normal distribution with correlation coefficient  $\rho$ .

*Proof*

We have seen from the proof of Proposition 1 that the FWER is controlled under  $H_1$  if  $(c'_1, c'_2) = (c_1, c_2)$ . Thus, we only need to consider the configuration  $\bar{H}_1 \cap H_2$ . As in the proof of Proposition 1, we first consider the FWER of  $\mathcal{P}$ , which can be written as  $P_1 + P_2$ , where

$$P_1 = P_{H_2}(X_1 > c_1, Y_1 > d_1) \text{ and } P_2 = P_{H_2}(X_1 \leq c_1, X_2 > c_2, Y_2 > d_2). \quad (17)$$

By conditioning on  $(X_1, Y_1) = (x_1, y_1)$  and by using the fact that under  $\bar{H}_1 \cap H_2$ ,  $(X_1 - \delta_1 \sqrt{n_1}, Y_1)$  is distributed as standard bivariate normal with correlation coefficient  $\rho$  and denoting its PDF  $\phi_2(x_1 - \delta_1 \sqrt{n_1}, y_1 | \rho)$  by  $f_{X_1, Y_1}(x_1, y_1)$ , we can write  $P_2$  as

$$\begin{aligned} P_2 &= \int_{-\infty}^{\infty} \int_{-\infty}^{c_1} P_{H_2}(X_2 > c_2, Y_2 > d_2 | x_1, y_1) f_{X_1, Y_1}(x_1, y_1) dx_1 dy_1 \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{c_1} P_{H_2}\left(X^{(2)} > \frac{c_2 - \sqrt{f}x_1}{\sqrt{1-f}}, Y^{(2)} > \frac{d_2 - \sqrt{f}y_1}{\sqrt{1-f}} \middle| x_1, y_1\right) f_{X_1, Y_1}(x_1, y_1) dx_1 dy_1 \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{c_1} P_{H_2}\left(Z_1 > \frac{c_2 - \sqrt{f}x_1}{\sqrt{1-f}} - \delta_1 \sqrt{n_2}, Z_2 > \frac{d_2 - \sqrt{f}y_1}{\sqrt{1-f}}\right) f_{X_1, Y_1}(x_1, y_1) dx_1 dy_1 \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{c_1} \Phi_2\left(\frac{\sqrt{f}x_1 - c_2}{\sqrt{1-f}} + \delta_1 \sqrt{n_2}, \frac{\sqrt{f}y_1 - d_2}{\sqrt{1-f}} \middle| \rho\right) \phi_2(x_1 - \delta_1 \sqrt{n_1}, y_1 | \rho) dx_1 dy_1. \quad (18) \end{aligned}$$

In the aforementioned derivation, we have used the fact that under  $\bar{H}_1 \cap H_2$ ,  $(X^{(2)} - \delta_1 \sqrt{n_2}, Y^{(2)})$  has a standard bivariate normal distribution with correlation coefficient  $\rho$ . The FWER of  $\mathcal{P}'_2$  under  $\bar{H}_1 \cap H_2$  can be analogously written as  $P'_1 + P'_2$ , where  $P'_1$  and  $P'_2$  are defined as in (17) but with  $(d_1, d_2)$  replaced by  $(d'_1, d'_2)$  and  $(X_2, Y_2)$  replaced by  $(X'_2, Y'_2)$ . Note that  $P'_1 = P_1$  because  $c'_1 = c_1$  and  $d'_1 = d_1$ . Furthermore,  $P'_2$  has the same integral expression (18) as for  $P_2$  but with  $d_2$  replaced by  $d'_2$  and  $n_2$  replaced by  $n'_2$  (which is fixed conditioned on  $X_1 = x_1$ ). Setting the integrands in the two integral expressions equal gives Equation (16).  $\square$

*Remark 2*

Note that  $d'_2$  depends on the observed  $(x_1, y_1)$ , and thus,  $\mathcal{P}'_2$  controls the FWER conditionally on  $(x_1, y_1)$ . However, because this holds for every  $(x_1, y_1)$ , it also holds unconditionally.

*Remark 3*

Equation (16) involves two unknown parameters,  $\rho$  and  $\delta_1$ . It is not clear how to use the upper confidence limit  $\rho^*$  in place of  $\rho$  because  $\rho^*$  and its associated confidence level  $1 - \varepsilon$  were determined in part I to optimize the  $(d_1, d_2)$  boundary. There is no analog of that problem here. So, we used the sample estimate  $r$  for  $\rho$  and similarly the sample estimate  $\hat{\delta}_1 = x_1 / \sqrt{n_1}$  for  $\delta_1$ . On the other hand, for choosing the unadjusted second stage critical boundary for the secondary endpoint,  $d_2$ , we employed the confidence limit method proposed in part I and used  $\rho^*$  for the unknown  $\rho$ .

## 5. Procedures based on sufficient statistics

We consider two adaptive group sequential procedures,  $\mathcal{P}''_1$  and  $\mathcal{P}''_2$ , based on the sufficient statistics  $(X_1, X''_2)$  and  $(Y_1, Y''_2)$  (where  $X''_2$  and  $Y''_2$  are defined in (8) and (9)). These are used in conjunction with critical boundaries  $(c''_1, c''_2)$  and  $(d''_1, d''_2)$  (different for  $\mathcal{P}''_1$  and  $\mathcal{P}''_2$ ), which are determined in the following text to control the FWER. First, we consider procedure  $\mathcal{P}''_1$ . In Proposition 3, we show that the critical boundaries  $(c''_1, c''_2)$  and  $(d''_1, d''_2)$  can be chosen so that  $\mathcal{P}''_1$  makes the same decisions as any GSP based on the CHW statistics that controls the FWER, for example,  $\mathcal{P}'_1$  or  $\mathcal{P}'_2$ , by modifying their critical constants. Thus,  $\mathcal{P}''_1$  is strictly not a sufficient statistics-based procedure; it simply re-expresses the rejection rule in terms of the sufficient statistics by making the second stage critical constants functions of the first stage data.

### 5.1. Procedure $\mathcal{P}''_1$

*Proposition 3*

Let  $(c_1, c_2)$  and  $(d_1, d_2)$  be the critical boundaries of  $\mathcal{P}'_1$  (or  $\mathcal{P}'_2$ ) that control the FWER at level  $\alpha$ . Denote the critical boundaries of  $\mathcal{P}''_1$  by  $(c''_1, c''_2)$  and  $(d''_1, d''_2)$ . Then,  $\mathcal{P}'_1$  (or  $\mathcal{P}'_2$ ) and  $\mathcal{P}''_1$  make identical decisions (and hence  $\mathcal{P}''_1$  controls the FWER at level  $\alpha$ ) if we set  $c''_1 = c_1, d''_1 = d_1,$

$$c''_2 = \frac{1}{\sqrt{n_1 + n'_2}} \left[ \sqrt{\frac{n'_2}{n_2}} (c_2 \sqrt{n_1 + n_2} - \sqrt{n_1} x_1) + \sqrt{n_1} x_1 \right] \quad (19)$$

and

$$d_2'' = \frac{1}{\sqrt{n_1 + n_2'}} \left[ \sqrt{\frac{n_2'}{n_2}} (d_2 \sqrt{n_1 + n_2} - \sqrt{n_1} y_1) + \sqrt{n_1} y_1 \right]. \quad (20)$$

*Proof*

Procedures  $\mathcal{P}'_1$  and  $\mathcal{P}''_1$  are equivalent at the first stage because  $c_1'' = c_1$  and  $d_1'' = d_1$ . For the second stage, it is easy to check using (7) and (8) that

$$X'_2 > c_2 \iff X^{(2)'} > \frac{c_2 - \sqrt{f} x_1}{\sqrt{1-f}}$$

and

$$X''_2 > c_2'' \iff X^{(2)''} > \frac{c_2'' - \sqrt{f'} x_1}{\sqrt{1-f'}}.$$

Equating the right-hand sides of the aforementioned two inequalities (to get the same rejection regions) and solving for  $c_2''$  yields (19). Equation (20) for  $d_2''$  is obtained in the same way.  $\square$

*Remark 4*

Gao *et al.* [6] provided the formula (19) for  $c_2''$ . They showed that in fact  $c_2'' \leq c_2$  for  $x_1 \in [1.1, 2]$  for  $\alpha = 0.025$ . Thus, in this region, we can increase  $n_2$  to  $n_2'$  without adjusting the critical constant  $c_2$  and the resulting procedure would still be conservative.

### 5.2. Procedure $\mathcal{P}''_2$

As noted before,  $\mathcal{P}''_1$  is simply a re-expression of  $\mathcal{P}'_1$ . To gain the potential power advantage associated with sufficient statistics, we need to determine  $(c_2'', d_2'')$  unconditionally by integrating over all possible outcomes  $(x_1, y_1)$  and solving the equation obtained by equating the resulting FWER integral to  $\alpha$ . Here, it is assumed that  $(c_1, d_1)$  are prespecified. We denote the corresponding procedure by  $\mathcal{P}''_2$ .

First, consider the problem of FWER control under  $H_1$ . It is well known [2, 12] that the  $\alpha$ -level boundary  $(c_1, c_2)$  does not control the FWER when used in conjunction with the sufficient test statistics  $(X_1, X_2'')$ . We can determine  $c_2''$  unconditionally as follows. For any given  $c_1$ , the equation for FWER control under  $H_1$  can be written as

$$\text{FWER} = P_1 + P_2 = \alpha,$$

where  $P_1 = \Phi(-c_1)$  (here  $\Phi(\cdot)$  is the standard normal CDF) and

$$\begin{aligned} P_2 &= P_{H_1}(X_1 \leq c_1, X_2'' > c_2'') \\ &= \int_{-\infty}^{c_1} P_{H_1}(\sqrt{f'} X^{(1)} + \sqrt{1-f'} X^{(2)''} > c_2'' | X^{(1)} = x_1) \phi(x_1) dx_1 \\ &= \int_{-\infty}^{c_1} P_{H_1}\left(X^{(2)''} > \frac{c_2'' - \sqrt{f'} x_1}{\sqrt{1-f'}}\right) \phi(x_1) dx_1 \\ &= \int_{-\infty}^{c_1} \Phi\left(\frac{-c_2'' + \sqrt{f'} x_1}{\sqrt{1-f'}}\right) \phi(x_1) dx_1, \end{aligned} \quad (21)$$

where  $f' = n_1/(n_1 + n_2')$  is function of  $x_1$  because  $n_2'$  depends on CP and hence on  $x_1$ . We can solve for  $c_2''$  by setting

$$P_2 = \int_{-\infty}^{c_1} \Phi\left(\frac{-c_2'' + \sqrt{f'} x_1}{\sqrt{1-f'}}\right) \phi(x_1) dx_1 = \alpha - \Phi(-c_1). \quad (22)$$

Next, we can derive the equation for determining  $d_2''$  to control the FWER under  $H_2$  as follows. First,

$$\text{FWER} = P_1 + P_2 = \alpha,$$

where

$$P_1 = P_{H_2}(X_1 > c_1, Y_1 > d_1) = \Phi_2(-c_1 + \delta_1 \sqrt{n_1}, -d_1 | \rho) \tag{23}$$

is a function of specified  $(c_1, d_1)$  and hence can be evaluated given  $n_1, \delta_1,$  and  $\rho$ . Next,

$$\begin{aligned} P_2 &= P_{H_2}(X_1 \leq c_1, X_2'' > c_2'', Y_2'' > d_2'') \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{c_1} P_{H_2} \left( \sqrt{f'} X^{(1)} + \sqrt{1-f'} X^{(2)} > c_2'', \sqrt{f'} Y^{(1)} \right. \\ &\quad \left. + \sqrt{1-f'} Y^{(2)} > d_2'' \mid X^{(1)} = x_1, Y^{(1)} = y_1 \right) \phi_2(x_1, y_1 | \rho) dx_1 dy_1 \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{c_1} P_{H_2} \left( X^{(2)} > \frac{c_2'' - \sqrt{f'} x_1}{\sqrt{1-f'}}, Y^{(2)} > \frac{d_2'' - \sqrt{f'} y_1}{\sqrt{1-f'}} \right) \phi_2(x_1, y_1 | \rho) dx_1 dy_1 \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{c_1} \Phi_2 \left( \frac{-c_2'' + \sqrt{f'} x_1 + \delta_1 \sqrt{n_2'}}{\sqrt{1-f'}}, \frac{-d_2'' + \sqrt{f'} y_1}{\sqrt{1-f'}} \mid \rho \right) \phi_2(x_1, y_1 | \rho) dx_1 dy_1. \end{aligned} \tag{24}$$

Having earlier obtained  $c_2''$  from (22), we can solve for  $d_2''$  from the aforementioned equation by setting (24) equal to  $\alpha - P_1$ , where  $P_1$  is given by (23).

### 6. Numerical evaluation of critical boundaries

Evaluation of the integrals in the aforementioned equations is complicated by the fact that  $f' = n_1/(n_1 + n_2')$  is a function of CP (because  $n_2'$  is) and hence of  $x_1$ . Numerical integration is more difficult for the Mehta–Pocock rule (12) than for the fixed increase rule, because in the former case,  $f'$  varies continuously with  $x_1$ , whereas in the latter,  $f'$  is a step function of  $x_1$ .

Computation of  $c_2''$  is relatively straightforward because Equation (21) does not involve any unknown parameters. However, computation of  $d_2''$  is more complicated because both  $P_1$  given by (23) and  $P_2$  given by (24) depend on the unknown parameters  $\delta_1$  and  $\rho$ . We will use the estimates  $\hat{\delta}_1 = x_1/\sqrt{n_1}$  and  $\hat{\rho} = r$  as proposed in Remark 3. Note that this makes  $d_2''$  conditional on the first stage data. However, as argued before, FWER is unconditionally controlled because it is controlled for every observed  $(x_1, y_1, r)$ .

Table I gives the computed values of  $(c_2'', d_2'')$  for the OF1–PO2 boundary ([13] boundary for the primary endpoint and the [14] boundary for the secondary endpoint) for  $\alpha = .05, \Delta_1 = 2.0, \rho = 0.5,$  and  $\gamma = 2, 4$  (typically, sample size increases would be limited to a factor of 2, but we also considered a factor of 4 to study any patterns in the computed values). We considered three promising zones: [30%, 70%], [40%, 80%], and [50%, 90%]. For the OF1 boundary, without sample size re-estimation, we have  $c_1 = 2.373$  and  $c_2 = 1.678$ . Similarly, for the PO2 boundary, we have  $d_1 = d_2 = 1.876$ . Note that  $c_2''$  values are computed constants because they do not depend on the estimates of any unknown parameters. However,  $d_2''$  values are random variables because they depend on  $(x_1, y_1, r)$ . The values reported in the table are averages obtained over 10,000 replications of  $(x_1, y_1, r)$ . It is seen that the  $c_2''$  values are larger and the  $d_2''$  values are smaller for  $\gamma = 2$  than for  $\gamma = 4$ . In either case,  $c_2'' < c_2$  and  $d_2'' < d_2$ . Thus, the critical boundaries of  $\mathcal{P}_2''$  are sharper than those of  $\mathcal{P}_1'$ . The  $c_2''$  values become smaller and the  $d_2''$  values become larger as the promising zone  $[CP_{\min}, CP_{\max}]$  shifts to the right, that is, the sample size adjustment is made for higher CP values.

Table I. $(c_2'', d_2'')$ values for the OF1–PO2 boundary using the fixed increase rule (13) for second stage sample size re-estimation ( $\alpha = .05, \Delta_1 = 2.0,$ and $\rho = 0.5$ ).		
$[CP_{\min}, CP_{\max}]$	$\gamma = 2$	$\gamma = 4$
[30%, 70%]	(1.6717, 1.7203)	(1.6575, 1.7400)
[40%, 80%]	(1.6624, 1.7231)	(1.6423, 1.7447)
[50%, 90%]	(1.6505, 1.7282)	(1.6222, 1.7534)



## 7. Power comparison

In this section, we compare the secondary powers (probability of rejecting false  $H_2$ ) of the following procedures via simulation:  $\mathcal{P}'_1$ , which uses separate  $\alpha$ -level boundaries  $(c_1, c_2)$  and  $(d_1, d_2)$ ,  $\mathcal{P}'_2$  using either the known  $\rho$  assumption (denoted as  $\mathcal{P}'_2(\rho)$ ) or the upper confidence limit  $\rho^*$  proposed in part I for dealing with unknown  $\rho$  (denoted as  $\mathcal{P}'_2(\rho^*)$ ), and  $\mathcal{P}''_2$  based on sufficient statistics. Note that because the first three procedures use the same primary test statistics  $(X_1, X'_2)$  and the same critical boundary  $(c_1, c_2)$ , their primary powers are identical.  $\mathcal{P}''_2$  will have a different primary power because it uses different test statistics  $(X_1, X''_2)$  and different critical boundary  $(c_1, c''_2)$ ; however, we did not study it in this paper.

We considered the same six combinations of different parameters as in Table I. Throughout, we kept the following quantities fixed: primary–secondary boundary combination: OF1–PO2,  $\alpha = .05$ ,  $n_1 = n_2 = 50$ ,  $\Delta_1 = 2.0$ ,  $\rho = 0.5$  and the number of replications per simulation run = 10,000. In each scenario, we varied  $\Delta_2$  from 1.0 to 4.0 in steps of 0.5. We present the results in Tables II–VII.

**Table II.** Secondary powers of  $\mathcal{P}'_1$ ,  $\mathcal{P}'_2(\rho^*)$ ,  $\mathcal{P}'_2(\rho)$ , and  $\mathcal{P}''_2$  using [30%, 70%] promising zone for conditional power and the fixed increase rule (13) with  $\gamma = 2$  for modified second stage sample size  $n'_2$  (OF1–PO2 boundary combination,  $\alpha = .05$ ,  $n_1 = n_2 = 50$ ,  $\Delta_1 = 2.0$ ,  $\rho = 0.5$ ).

$\Delta_2$	$\mathcal{P}'_1$	$\mathcal{P}'_2(\rho^*)$	$\mathcal{P}'_2(\rho)$	$\mathcal{P}''_2$
1.0	0.290	0.350	0.364	0.363
1.5	0.549	0.595	0.609	0.620
2.0	0.744	0.773	0.785	0.785
2.5	0.851	0.864	0.867	0.868
3.0	0.884	0.888	0.889	0.889
3.5	0.899	0.900	0.900	0.900
4.0	0.900	0.900	0.900	0.901

**Table III.** Secondary powers of  $\mathcal{P}'_1$ ,  $\mathcal{P}'_2(\rho^*)$ ,  $\mathcal{P}'_2(\rho)$ , and  $\mathcal{P}''_2$  using [40%, 80%] promising zone for conditional power and the fixed increase rule (13) with  $\gamma = 2$  for modified second stage sample size  $n'_2$  (OF1–PO2 boundary combination,  $\alpha = .05$ ,  $n_1 = n_2 = 50$ ,  $\Delta_1 = 2.0$ ,  $\rho = 0.5$ ).

$\Delta_2$	$\mathcal{P}'_1$	$\mathcal{P}'_2(\rho^*)$	$\mathcal{P}'_2(\rho)$	$\mathcal{P}''_2$
1.0	0.292	0.338	0.355	0.369
1.5	0.536	0.586	0.601	0.610
2.0	0.742	0.771	0.780	0.785
2.5	0.841	0.855	0.859	0.859
3.0	0.891	0.896	0.897	0.899
3.5	0.894	0.895	0.895	0.896
4.0	0.898	0.898	0.898	0.900

**Table IV.** Secondary powers of  $\mathcal{P}'_1$ ,  $\mathcal{P}'_2(\rho^*)$ ,  $\mathcal{P}'_2(\rho)$ , and  $\mathcal{P}''_2$  using [50%, 90%] promising zone for conditional power and the fixed increase rule (13) with  $\gamma = 2$  for modified second stage sample size  $n'_2$  (OF1–PO2 boundary combination,  $\alpha = .05$ ,  $n_1 = n_2 = 50$ ,  $\Delta_1 = 2.0$ ,  $\rho = 0.5$ ).

$\Delta_2$	$\mathcal{P}'_1$	$\mathcal{P}'_2(\rho^*)$	$\mathcal{P}'_2(\rho)$	$\mathcal{P}''_2$
1.0	0.300	0.350	0.364	0.374
1.5	0.545	0.596	0.611	0.618
2.0	0.744	0.775	0.782	0.788
2.5	0.843	0.856	0.859	0.860
3.0	0.887	0.892	0.893	0.894
3.5	0.900	0.901	0.902	0.903
4.0	0.900	0.900	0.900	0.903

**Table V.** Secondary powers of  $\mathcal{P}'_1$ ,  $\mathcal{P}'_2(\rho^*)$ ,  $\mathcal{P}'_2(\rho)$ , and  $\mathcal{P}''_2$  using [30%, 70%] promising zone for conditional power and the fixed increase rule (13) with  $\gamma = 4$  for modified second stage sample size  $n'_2$  (OF1–PO2 boundary combination,  $\alpha = .05$ ,  $n_1 = n_2 = 50$ ,  $\Delta_1 = 2.0$ ,  $\rho = 0.5$ ).

$\Delta_2$	$\mathcal{P}'_1$	$\mathcal{P}'_2(\rho^*)$	$\mathcal{P}'_2(\rho)$	$\mathcal{P}''_2$
1.0	0.317	0.364	0.380	0.397
1.5	0.567	0.609	0.623	0.633
2.0	0.759	0.786	0.794	0.796
2.5	0.852	0.864	0.868	0.868
3.0	0.892	0.898	0.899	0.900
3.5	0.904	0.906	0.906	0.909
4.0	0.906	0.906	0.906	0.908

**Table VI.** Secondary powers of  $\mathcal{P}'_1$ ,  $\mathcal{P}'_2(\rho^*)$ ,  $\mathcal{P}'_2(\rho)$ , and  $\mathcal{P}''_2$  using [40%, 80%] promising zone for conditional power and the fixed increase rule (13) with  $\gamma = 4$  for modified second stage sample size  $n'_2$  (OF1–PO2 boundary combination,  $\alpha = .05$ ,  $n_1 = n_2 = 50$ ,  $\Delta_1 = 2.0$ ,  $\rho = 0.5$ ).

$\Delta_2$	$\mathcal{P}'_1$	$\mathcal{P}'_2(\rho^*)$	$\mathcal{P}'_2(\rho)$	$\mathcal{P}''_2$
1.0	0.326	0.375	0.390	0.408
1.5	0.579	0.626	0.639	0.645
2.0	0.762	0.793	0.800	0.804
2.5	0.849	0.863	0.867	0.871
3.0	0.893	0.898	0.900	0.903
3.5	0.904	0.906	0.906	0.910
4.0	0.909	0.909	0.909	0.913

**Table VII.** Secondary powers of  $\mathcal{P}'_1$ ,  $\mathcal{P}'_2(\rho^*)$ ,  $\mathcal{P}'_2(\rho)$ , and  $\mathcal{P}''_2$  using [50%, 90%] promising zone for conditional power and the fixed increase rule (13) with  $\gamma = 4$  for modified second stage sample size  $n'_2$  (OF1–PO2 boundary combination,  $\alpha = .05$ ,  $n_1 = n_2 = 50$ ,  $\Delta_1 = 2.0$ ,  $\rho = 0.5$ ).

$\Delta_2$	$\mathcal{P}'_1$	$\mathcal{P}'_2(\rho^*)$	$\mathcal{P}'_2(\rho)$	$\mathcal{P}''_2$
1.0	0.344	0.393	0.407	0.426
1.5	0.594	0.637	0.651	0.656
2.0	0.765	0.792	0.800	0.803
2.5	0.858	0.869	0.871	0.876
3.0	0.890	0.894	0.894	0.903
3.5	0.902	0.903	0.903	0.910
4.0	0.905	0.905	0.905	0.913

For each simulation run, we computed the percentage power gain achieved by  $\mathcal{P}'_2(\rho^*)$  over  $\mathcal{P}'_1$  as a fraction of the maximum achievable power gain of the gold standard procedure  $\mathcal{P}'_2(\rho)$  for known  $\rho$  over  $\mathcal{P}'_1$  (which assumes  $\rho = 1$ ), as defined in the following:

$$\text{Power Gain (\%)} = \frac{\text{Power}(\rho = \rho^*) - \text{Power}(\rho = 1)}{\text{Power}(\text{Known } \rho) - \text{Power}(\rho = 1)} \times 100. \tag{25}$$

For all scenarios, these power gains range between 70% and 80%. Procedure  $\mathcal{P}''_2$  based on the sufficient statistics generally has the highest power, even higher than that of the gold standard procedure  $\mathcal{P}'_2(\rho)$ , demonstrating the power advantage of sufficient statistics. However, one must weigh this power advantage of  $\mathcal{P}''_2$  against the difficulty of computing its critical constants ( $c''_2, d''_2$ ).

### 8. Extension to two samples

The basic setup and notation are the same as defined in Section 7 of part I. The stagewise test statistics are given by

$$X^{(k)} = \frac{\bar{U}_{1\cdot k} - \bar{U}_{2\cdot k}}{\sigma_1 \sqrt{1/n_{1k} + 1/n_{2k}}} \text{ and } Y^{(k)} = \frac{\bar{V}_{1\cdot k} - \bar{V}_{2\cdot k}}{\sigma_2 \sqrt{1/n_{1k} + 1/n_{2k}}} \quad (k = 1, 2), \tag{26}$$

where  $\bar{U}_{i\cdot k}$  and  $\bar{V}_{i\cdot k}$  are the sample means of the observations  $U_{ijk}$  and  $V_{ijk}$  averaged over patients  $j = 1, \dots, n_{ik}$ , respectively. Next, denote by  $\bar{U}_{i\cdot}$  and  $\bar{V}_{i\cdot}$  the overall sample means of the primary and secondary endpoints data for the  $i$ th group and  $n_{i\cdot} = n_{i1} + n_{i2}$  ( $i = 1, 2$ ). The cumulative test statistics  $(X_1, Y_1)$  and  $(X_2, Y_2)$  for the first and second stages are as defined in (11) and (12), respectively, of part I. Note that we can express  $X_2, Y_2$  as follows:

$$X_2 = \frac{n_{11}\sqrt{\frac{n_{2\cdot}}{n_{1\cdot}}}\bar{U}_{1\cdot 1} - n_{21}\sqrt{\frac{n_{1\cdot}}{n_{2\cdot}}}\bar{U}_{2\cdot 1}}{\sigma_1\sqrt{n_{1\cdot} + n_{2\cdot}}} + \frac{n_{12}\sqrt{\frac{n_{2\cdot}}{n_{1\cdot}}}\bar{U}_{1\cdot 2} - n_{22}\sqrt{\frac{n_{1\cdot}}{n_{2\cdot}}}\bar{U}_{2\cdot 2}}{\sigma_1\sqrt{n_{1\cdot} + n_{2\cdot}}}$$

and

$$Y_2 = \frac{n_{11}\sqrt{\frac{n_{2\cdot}}{n_{1\cdot}}}\bar{V}_{1\cdot 1} - n_{21}\sqrt{\frac{n_{1\cdot}}{n_{2\cdot}}}\bar{V}_{2\cdot 1}}{\sigma_2\sqrt{n_{1\cdot} + n_{2\cdot}}} + \frac{n_{12}\sqrt{\frac{n_{2\cdot}}{n_{1\cdot}}}\bar{V}_{1\cdot 2} - n_{22}\sqrt{\frac{n_{1\cdot}}{n_{2\cdot}}}\bar{V}_{2\cdot 2}}{\sigma_2\sqrt{n_{1\cdot} + n_{2\cdot}}}.$$

These can be expressed as linear combinations of the stagewise test statistics  $(X^{(k)}, Y^{(k)})$  for  $k = 1, 2$  only if  $n_{11} = n_{21}$  and  $n_{12} = n_{22}$ , that is, if the treatment and control groups have a balanced sample size allocation in both stages. From now on, we will assume this condition (which may not hold exactly when the patients are randomized to the treatment and control groups). Then, it is easy to show that  $X_2, Y_2$  can be written as follows:

$$X_2 = \sqrt{f}X^{(1)} + \sqrt{1-f}X^{(2)} \text{ and } Y_2 = \sqrt{f}Y^{(1)} + \sqrt{1-f}Y^{(2)}, \quad (27)$$

where

$$f = \frac{n_{11} + n_{21}}{n_{1\cdot} + n_{2\cdot}} = \frac{n_{\cdot 1}}{n_{\cdot 1} + n_{\cdot 2}} \quad (28)$$

is the fraction of the total sample size allocated to the first stage, which is analogous to the information fraction  $f = n_1/(n_1 + n_2)$  in the one-sample case.

Next, consider the case where the second stage total sample size  $n_{\cdot 2} = n_{12} + n_{22}$  is re-estimated to  $n'_{\cdot 2} = n'_{12} + n'_{22}$ . Let  $\bar{U}'_{i\cdot 2}$  and  $\bar{V}'_{i\cdot 2}$  ( $i = 1, 2$ ) be the sample means of the second stage data with re-estimated sample sizes. The corresponding second stage standardized test statistics  $X'^{(2)}$  and  $Y'^{(2)}$  are defined as in (26) for  $k = 2$  but with all quantities in the formulae replaced by their primed analogs. Once again assuming a balanced sample size allocation with  $n_{11} = n_{21}$  and  $n'_{12} = n'_{22}$ , the second stage CHW statistics are given by (7) with  $X^{(1)}, X'^{(2)}, Y^{(1)}, Y'^{(2)}$  for the two-sample problem as defined previously and  $f$  given by (28). Similarly, the sufficient statistics are given by (8) and (9) with

$$f' = \frac{n_{\cdot 1}}{n_{\cdot 1} + n'_{\cdot 2}}.$$

An alternative and a more direct way of defining sufficient statistics is

$$X'_2 = \frac{\bar{U}_{1\cdot\cdot} - \bar{U}_{2\cdot\cdot}}{\sigma_2\sqrt{1/n_{1\cdot} + 1/n_{2\cdot}}} \text{ and } Y'_2 = \frac{\bar{V}_{1\cdot\cdot} - \bar{V}_{2\cdot\cdot}}{\sigma_2\sqrt{1/n_{1\cdot} + 1/n_{2\cdot}}}. \quad (29)$$

This definition does not require the assumption of the balanced sample size allocation. We will use this definition in the example in the following text.

We have assumed that the standard deviations  $\sigma_1$  and  $\sigma_2$  are known, but in practice, they must be estimated. The first stage pooled (from the treatment and the control groups) estimate of  $\sigma_1$  is given by

$$\hat{\sigma}_1^{(1)} = \sqrt{\frac{\sum_{i=1}^2 \sum_{j=1}^{n_{i1}} (U_{ij1} - \bar{U}_{i\cdot 1})^2}{n_{11} + n_{21} - 2}}$$

with an analogous expression for  $\hat{\sigma}_2^{(1)}$ . These estimates will be used to calculate  $X^{(1)}$  and  $Y^{(1)}$ . The second stage pooled estimates,  $\hat{\sigma}_1^{(2)}$  and  $\hat{\sigma}_2^{(2)}$ , have similar expressions except the sample sizes  $n_{i2}$  are changed to the re-estimated sample sizes  $n'_{i2}$ . Because  $X'^{(2)}$  and  $Y'^{(2)}$  must be based only on the second

stage data, these second stage pooled estimates are used in their calculation and not the overall pooled estimates  $\hat{\sigma}_1$  and  $\hat{\sigma}_2$ , where

$$\hat{\sigma}_1 = \sqrt{\frac{\sum_{i=1}^2 \sum_{j=1}^{n_{i1}} (U_{ij1} - \bar{U}_{i..})^2 + \sum_{i=1}^2 \sum_{j=1}^{n'_{i2}} (U_{ij2} - \bar{U}_{i..})^2}{n_{.1} + n'_{.2} - 2}}$$

with an analogous expression for  $\hat{\sigma}_2$ . We use these overall pooled estimates in the calculation of the sufficient statistics (29) in the example in the following text.

### 9. Example

We will use the same ISOLDE trial example from part I. As discussed there, we will use the rate of decline in forced expiratory volume at 1 s (FEV1) as the primary endpoint and the rate of decline in forced vital capacity(FVC) as the secondary endpoint where the rates of decline are computed by dividing the difference between the final measurement (the timing of which varies from patient to patient depending upon how many visits they completed) and the baseline measurement (at randomization) by the period (in months) between the two measurements.

To allow for sample size re-estimation with increased sample size at the second stage, we assume that the original trial was planned with a total of 300 patients with  $n_{.1} = n_{.2} = 150$  patients in each stage (randomized between the treatment and the placebo). This design corresponds to primary power of 90% to detect a clinically significant difference of  $\delta_1 = (\mu_1 - \mu_2)/\sigma_1 = 0.53$  or approximately 0.50. The sample size re-estimation rule is as follows: if CP at the interim look is in the promising zone [50%, 90%], then we would use  $n'_{.2} = 2n_{.2} = 300$  patients bringing the total number of patients to 450. We further assume that the OF1–PO2 boundary combination with  $\alpha = 0.05$  is used for the primary and secondary endpoints. The corresponding OF1 critical values are  $c_1 = 2.7959, c_2 = 1.9770$  and the PO2 critical values are  $d_1 = d_2 = 2.0661$ .

Of the first 150 patients,  $n_{11} = 69$  were randomized to the treatment and  $n_{21} = 81$  to the control. At the interim look, we have the following summary statistics:

$$\bar{U}_{1.1} = -0.0029, \bar{U}_{2.1} = -0.0070, \bar{V}_{1.1} = -0.0091, \bar{V}_{2.1} = -0.0142,$$

$$\hat{\sigma}_1^{(1)} = 0.0143, \hat{\sigma}_2^{(1)} = 0.0304, r = 0.6661.$$

From these summary statistics, we can calculate  $X^{(1)} = 1.7783$  and  $Y^{(1)} = 1.0151$ . Because  $X^{(1)} < c_1$ , sampling continues to the second stage. To determine if the second stage sample size needs to be re-estimated, we calculate the CP. We have  $\hat{\delta}_1 = (\bar{U}_{1.1} - \bar{U}_{2.1})/\hat{\sigma}_1^{(1)} = (-0.0029 + 0.0070)/0.0143 = 0.2913$ . CP is given by the formula (10) with  $n_2$  replaced by its two-sample analog  $(n_{12}n_{22})/(n_{12} + n_{22})$ . At the interim look,  $n_{12}$  and  $n_{22}$  are as yet unobserved, so we take  $n_{12} = n_{22} = 150/2 = 75$ , which gives  $(n_{12}n_{22})/(n_{12} + n_{22}) = 37.5$ . Also,  $f = 0.5$ . Substituting these values in (10), we get

$$CP = 1 - \Phi\left(\frac{1.9770 - 1.7783\sqrt{0.5}}{\sqrt{1 - 0.5}} - 0.2913\sqrt{37.5}\right) = 1 - \Phi(-0.7659) = 0.7782.$$

Because CP falls in the promising zone, we increase the second stage sample from 150 to 300.

It would be instructive to compare this sample size with what one gets using the Gao–Ware–Mehta formula (11). Modified for the two-sample setup, that formula becomes

$$n'_{.2} = \frac{4}{\delta_1^2} \left\{ \frac{1}{n_{.2}} [c_2 \sqrt{n_{.1} + n_{.2}} - x_1 \sqrt{n_{.1}}] + z_\beta \right\}^2.$$

Substituting  $\hat{\delta}_1 = 0.2913, n_{.1} = n_{.2} = 150, c_2 = 1.9770, x_1 = 1.7783$ , and  $z_\beta = 1.282$  for  $1 - \beta = 0.90$ , we get  $n'_{.2} = 249.3$  or 250.

In the actual data if we consider the next 300 patients, then we find that  $n_{12} = 161$  were randomized to the treatment and  $n_{22} = 139$  to the control resulting in a total of  $n_{1.} = 230$  patients on the treatment and  $n_{2.} = 220$  on the control. The second stage summary statistics are as follows:

$$\bar{U}_{1.2} = -0.0043, \bar{U}_{2.2} = -0.0084, \bar{V}_{1.2} = -0.0070, \bar{V}_{2.2} = -0.0135, \hat{\sigma}_1^{(2)} = 0.0119, \hat{\sigma}_2^{(2)} = 0.0278.$$

From these summary statistics, we calculate  $X'(2) = 2.9195$  and  $Y'(2) = 2.0069$ .

The CHW statistics equal

$$X'_2 = \sqrt{f}X^{(1)} + \sqrt{1-f}X^{(2)} = \sqrt{1/2}(1.7783) + \sqrt{1/2}(2.9195) = 3.3218$$

and

$$Y'_2 = \sqrt{f}Y^{(1)} + \sqrt{1-f}Y^{(2)} = \sqrt{1/2}(1.0151) + \sqrt{1/2}(2.0069) = 2.1369.$$

$\mathcal{P}'_1$  compares  $X'_2$  and  $Y'_2$  with  $c_2 = 1.9770$  and  $d_2 = 2.0661$ , respectively. Thus,  $\mathcal{P}'_1$  rejects both  $H_1$  and  $H_2$ .  $\mathcal{P}'_2$  compares  $X'_2$  with  $c_2 = 1.9770$  but  $Y'_2$  with  $d'_2 = 2.0665$ , which is determined from (16) given the observed values of  $(X_1, Y_1)$ ,  $\rho$  estimated by  $r = 0.6661$  and  $\delta_1$  estimated by  $\hat{\delta}_1 = 0.2913$ . So, we obtain the same result.

Next, we will apply  $\mathcal{P}''_2$  based on the sufficient statistics. The overall sample means and the standard deviations are

$$\bar{U}_{1..} = -0.0039, \bar{U}_{2..} = -0.0079, \bar{V}_{1..} = -0.0076, \bar{V}_{2..} = -0.0137, \hat{\sigma}_1 = 0.0127, \hat{\sigma}_2 = 0.0287.$$

Using (29), we can calculate the sufficient statistics as

$$X''_2 = \frac{-0.0039 + 0.0079}{0.0127\sqrt{1/230 + 1/220}} = 3.3398$$

and

$$Y''_2 = \frac{-0.0076 + 0.0137}{0.0287\sqrt{1/230 + 1/220}} = 2.2538.$$

To apply  $\mathcal{P}''_2$ , we calculated  $c''_2 = 1.9497$  from (23) and  $d''_2 = 2.0555$  from (24) given  $(X_1, Y_1) = (1.7783, 1.0151)$ ,  $\rho$  estimated by  $r = 0.6661$  and  $\delta_1$  estimated by  $\hat{\delta}_1 = 0.2913$ . These values are slightly smaller than  $c_2$  and  $d_2$ , respectively. Once again,  $X''_2 > c''_2$  and  $Y''_2 > d''_2$ , so we get the same result.

## 10. Discussion

Although  $\mathcal{P}''_2$  based on sufficient statistics is the most powerful among the procedures that were compared, there are some practical drawbacks associated with it. First, computation of its critical constants ( $c''_2, d''_2$ ) is rather complicated. Second, and more important, it requires that ( $c''_2, d''_2$ ) be prespecified, which makes the sample size increase binding if the CP falls in the promising zone; otherwise, the stopping boundary will not be valid. This ties the hands of the Data Monitoring Committee, which may not want to increase the sample size because of other reasons such as slow accrual rate or excessive toxicity. On the other hand, procedure  $\mathcal{P}'_2(\rho^*)$  gives the flexibility to change the plans about sample size adaptation without the risk of inflating the type I error. These practical considerations must be weighed against the fact that the procedure  $\mathcal{P}''_2$  has the higher secondary power. Thus, although the CHW statistics are less efficient, they are more practically applicable because of their flexibility.

The problem of binary data was considered in the Discussion section of part I. It was noted there that there are difficulties in extending the normal data results to binary data especially in the two-sample case because of the dependence of the correlation coefficient between the primary and secondary endpoints on the corresponding success probabilities for the treatment and control groups, which must be equal (i.e.,  $H_1$  and  $H_2$  must be true) for the assumption of the common correlation coefficient for the two groups to be valid.

## Acknowledgements

The authors thank the referees for their useful suggestions, Dr. Lingyun Liu for providing the proof of Proposition 1, and Dr. Stuart Pocock for providing the data used in the ISOLDE trial.

## References

1. Fisher LD. Self-designing clinical trials. *Statistics in Medicine* 1998; **17**:1551–1562.
2. Cui L, Hung HMJ, Wang SJ. Modification of sample size in group sequential clinical trials. *Biometrics* 1999; **55**:853–857.
3. Jennison C, Turnbull BW. Mid-course sample size modification in clinical trials based on the observed treatment effect. *Statistics in Medicine* 2003; **22**:971–993.
4. Jennison C, Turnbull BW. Adaptive and nonadaptive group sequential tests. *Biometrika* 2006; **93**:1–21.
5. Jennison C, Turnbull BW. Adaptive seamless designs: selection and prospective testing of hypotheses. *Journal of Biopharmaceutical Statistics* 2007; **17**:1135–1161.
6. Gao P, Ware JH, Mehta CR. Sample size re-estimation for adaptive sequential design. *Journal of Biopharmaceutical Statistics* 2008; **18**:1184–1196.
7. Mehta CR, Pocock SJ. Adaptive increase in sample size when interim results are promising: a practical guide with examples. *Statistics in Medicine* 2010. published online DOI: 10.1002/sim.4102.
8. Lan KKG, DeMets DL. Discrete sequential boundaries for clinical trials. *Biometrika* 1983; **70**:659–663.
9. Tamhane AC, Mehta CR, Liu L. Testing a primary and secondary endpoint in a group sequential design. *Biometrics* 2010; **66**:1174–1184.
10. Tamhane AC, Wu Y, Mehta CR. An adaptive two-stage group sequential procedure for testing a primary and a secondary endpoint using the sample estimate of the unknown correlation between the endpoints 2012, *Statistics in Medicine*. published online DOI: 10.1002/sim.5372.
11. Liu L. Personal communication, 2011.
12. Proschan MA, Hunsberger SA. Designed extension of studies based on conditional power. *Biometrics* 1995; **51**:1315–1324.
13. O'Brien PC, Fleming TR. A multiple testing procedure for clinical trials. *Biometrics* 1979; **35**:549–556.
14. Pocock SJ. Group sequential methods in the design and analysis of clinical trials. *Biometrika* 1977; **64**:191–199.