Unknown Correlation Coefficient Between a Primary and a Secondary Endpoint in a Two-stage Group Sequential Design

Yi Wu* Ajit C. Tamhane† Cyrus R. Mehta‡

Abstract

We consider a two-stage group sequential procedure (GSP) with a primary and a secondary endpoint where the secondary endpoint is tested only if the primary endpoint shows significance. We assume that the two endpoints follow a bivariate normal distribution with unknown correlation coefficient ρ. Setting $\rho = 1$ will provide the most conservative critical boundary (Tamhane, Mehta and Liu 2010). However, replacing ρ with its sample estimate r will cause familywise error rate (FWER) inflation when r is less than ρ. We have developed an approach to control the overall FWER with adjusted second stage critical boundary utilizing the upper confidence limit of ρ. We show that the power gain for rejecting the secondary hypothesis of the new approach over the conservative method is significant.

Key Words: Multiple endpoints; Group Sequential Procedures; Gatekeeping procedures;

1. Introduction

Pocock (1977) proposed a group sequential procedure (GSP) with constant nominal significant levels. O'Brien and Fleming (1979) proposed another GSP using lower nominal significant levels at earlier stages. Their wide early boundaries are often preferred in practice since problems in data quality may arise at the beginning of a study. Lan and DeMets (1983) allowed group sequential designs to be extended to a more general setting in which the error spending function and decision times are allowed to be flexible. However, most of their work relates to a single endpoint. Jennison and Turnbull (2000) provided a comprehensive reference on this topic.

Tamhane, Mehta and Liu (2010) proposed a two-stage one-sided group sequential procedure (GSP) with a primary and a secondary endpoint where the primary endpoint is a gatekeeper for the secondary endpoint. They assume that the two endpoints follow a bivariate normal distribution with unknown correlation coefficient ρ. In their paper, the first and second stage critical boundaries are provided to control the FWER in two circumstances: (i) ρ is known or (ii) $\rho = 1$, which is the least favorable case. However, neither assumption is very practical. The $\rho = 1$ assumption is too conservative resulting in loss of power, while the known ρ assumption, although it yields a more powerful GSP, is never true in practice. In this paper, we show how to use sample estimate of ρ to adjust the secondary critical boundary in order to control the FWER.

The paper is organized as follows. Section 2 defines the notation and gives a brief review of the Tamhane et al. (2010) paper. Section 3 presents the confidence limit method to deal with unknown correlation. Section 4 gives the results of the secondary power comparisons of the proposed method with the most conservative method that assumes $\rho = 1$ and the least conservative method that assumes that the true ρ is known. Section 5 gives a clinical example and some concluding remarks are given in Section 6.

*Department of Statistics, Northwestern University, Evanston, IL 60208

†Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL 60208

‡Cytel Corporation and Department of Biostatistics, Harvard School of Public Health, Cambridge, MA 02139

2. Notation and Background

Assume a two-stage GSP with sample sizes n_1 and n_2. Let (X_{ij}, Y_{ij}) be i.i.d. bivariate normal observations on the primary and secondary endpoints for the jth patient in the ith group $(i = 1, 2, j = 1, \ldots, n_i)$ 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 the sake of simplicity we will assume that σ_1 and σ_2 are known. 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 iff H_1 is rejected; otherwise H_2 is accepted without a test.

The test statistics at the first stage are defined as

$$X_1 = \frac{\sum_{j=1}^{n_1} X_{1j}}{\sigma_1\sqrt{n_1}}, Y_1 = \frac{\sum_{j=1}^{n_1} Y_{1j}}{\sigma_2\sqrt{n_1}}, \tag{1}$$

and those at the second stage are defined as

$$X_2 = \frac{\sum_{j=1}^{n_1} X_{1j} + \sum_{j=1}^{n_2} X_{2j}}{\sigma_1\sqrt{n_1+n_2}}, Y_2 = \frac{\sum_{j=1}^{n_1} Y_{1j} + \sum_{j=1}^{n_2} Y_{2j}}{\sigma_2\sqrt{n_1+n_2}}. \tag{2}$$

Denote the primary critical boundary for (X_1, X_2) by (c_1, c_2) and the secondary critical boundary for (Y_1, Y_2) by (d_1, d_2). The GSP, denoted by $\mathcal{P}$, operates as follows.

Stage 1: Take n_1 observations, $(X_{1j}, Y_{1j}), j = 1, \ldots, n_1$, 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.

Stage 2: Take n_2 observations, $(X_{2j}, Y_{2j}), j = 1, \ldots, n_2$, 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.

The critical boundaries of $\mathcal{P}$ must be determined to control the FWER, defined as

$$\text{FWER} = P(\text{Reject at least one true } H_i \ (i = 1, 2)), \tag{3}$$

at level $\leq \alpha$ for specified α. Throughout we will assume that (c_1, c_2) and (d_1, d_2) are chosen so that $\mathcal{P}$ satisfies this FWER requirement with known true ρ.

It is easy to see that to control the FWER under H_1, (c_1, c_2) must be an α-level boundary, e.g., the O'Brien-Fleming (OF) (1979) boundary or the Pocock (PO) (1977) boundary. Under H_2, FWER is a function of $\Delta_1 = \delta_1\sqrt{n_1}$ and ρ (denoted by $\text{FWER}(\Delta_1, \rho)$). To control FWER in this case, for given (c_1, c_2), we need to determine (d_1, d_2) so that $\max_{\Delta_1,\rho} \text{FWER}(\Delta_1, \rho) \leq \alpha$.

It was shown in Tamhane et al. (2010) that $\max_{\Delta_1} \text{FWER}(\Delta_1, \rho)$ is an increasing function of ρ and the overall maximum of $\text{FWER}(\Delta_1, \rho))$ occurs when $\rho = 1$ and $\Delta_1 = \Delta_1^*$ where Δ_1^* depends on (c_1, c_2) and (d_1, d_2).

3. Confidence Limit Method

Since $\max_{\Delta_1} \text{FWER}(\Delta_1, \rho)$ is an increasing function of ρ, the (d_1, d_2) boundary becomes more conservative as the assumed value of ρ gets larger, $\rho = 1$ being the least favorable of ρ. An important practical problem is how to choose this boundary when ρ is unknown by using the sample correlation coefficient r from the first stage data. Note that we are assuming that, as mentioned before, (c_1, c_2) is an *a priori* specified α-level boundary, but

(d_1, d_2) can be adaptively determined as long as they are not functions of (X_1, Y_1). In the present case they will be functions of r, which is independent of (X_1, Y_1).

If (d_1, d_2) are determined simply by substituting the sample r for the unknown ρ then the FWER will be overestimated if $r < \rho$ and underestimated if $r > \rho$. Thus, even though the FWER may be close to the nominal α on the average, in a significant proportion of cases it may exceed α. Figure 1 shows the plot of simulated proportion of times the FWER exceeds the nominal value $\alpha = 0.05$ as a function of Δ_1 when the true $\rho = 0.5$. We see that when $\Delta_1 \approx 1.5$, the FWER exceeds $\alpha = 0.05$ in about 50% of the cases. Therefore we should not simply substitute r for the true ρ. However, given that $\max_{\Delta_1}$ FWER(Δ_1, ρ) is an increasing function of ρ, we can determine a conservative boundary (d_1, d_2) based on an upper confidence limit on ρ.

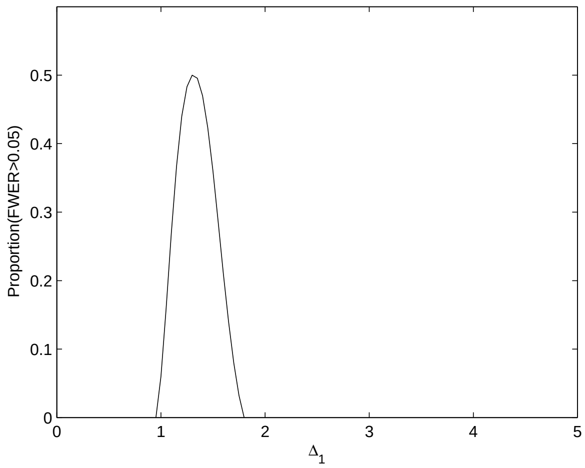

Figure 1: Proportion of simulation runs in which FWER > 0.05 if r is used as the true ρ $(\rho = 0.5)$

Let ρ^* be a $100(1-\varepsilon)\%$ upper confidence limit on ρ, i.e., $P(\rho \leq \rho^*) = 1 - \varepsilon$. In our calculations we used the following confidence limit based on Fisher's arctan hyperbolic transformation:

$$\rho^* = \frac{e^{2u} - 1}{e^{2u} + 1} \text{ where } u = \frac{1}{2} \ln\left(\frac{1+r}{1-r}\right) + \frac{z_\varepsilon}{\sqrt{n_1 - 3}}, \tag{4}$$

where z_ε is the $100(1-\varepsilon)\%$ percentile of the standard normal distribution. Then, using the property that $\max_{\Delta_1}$ FWER(Δ_1, ρ) is an increasing function of ρ, the overall maximum FWER can be written as follows.

Let $\Delta_1^*(\rho)$ be the value of Δ_1 that maximizes FWER(Δ_1, ρ) for fixed ρ. Then

$$\begin{aligned} & \max_{\{\Delta_1, \rho\}} \text{FWER}(\Delta_1, \rho) \\ = \; & \max_{\{\rho \leq \rho^*\}} \text{FWER}(\Delta_1^*(\rho), \rho) \times P(\rho \leq \rho^*) + \max_{\{\rho > \rho^*\}} \text{FWER}(\Delta_1^*(\rho), \rho) \times P(\rho > \rho^*) \\ = \; & \text{FWER}(\Delta_1^*(\rho^*), \rho^*) \times (1 - \varepsilon) + \text{FWER}(\Delta_1^*(1), 1) \times \varepsilon, \end{aligned} \tag{5}$$

where we have used the fact that the maximum of FWER$(\Delta_1^*(\rho), \rho)$ over a given range of ρ occurs at the upper limit of that range.

We want to determine the sharpest possible (d_1, d_2) (so as to maximize the secondary power) subject to the above expression being $\leq \alpha$. This problem can be solved numerically on a computer as follows. For fixed (d_1, d_2) we can evaluate FWER$(\Delta_1^*(\rho^*), \rho^*) = \alpha'$ (say) and FWER$(\Delta_1^*(1), 1) = \alpha''$ (say) where $\alpha' < \alpha''$ are functions of (d_1, d_2). Generally, (d_1, d_2) are parameterized through some common d, e.g., for the OF boundary, $d_1 = d/\sqrt{t_1}, d_2 = d$ and for the PO boundary, $d_1 = d_2 = d$.

We optimize the confidence level $1 - \varepsilon$ so as to minimize the common d and thus maximize the secondary power. First note that if $1 - \varepsilon$ is increased then ρ^* increases causing α' to increase while α'' remains fixed. Also, the weight $1 - \varepsilon$ on α' increases while the weight ε on α'' decreases. The net result is that as $1 - \varepsilon$ increases, the overall $\max$ FWER first decreases (since the weight ε on α'', which is greater than α', decreases) and then increases. We should then choose, for each given n_1 and r, the value of $1 - \varepsilon$ that minimizes the overall $\max$ FWER and choose d to make this minimax FWER equal to α (which will in turn maximize the secondary power for given α).

To calculate the optimum (d_1, d_2) boundary for given sample correlation coefficient r, the primary boundary (c_1, c_2), and the sample size $n_1 = n_2 = n$, we considered four cases for $\alpha = 0.05$: the (c_1, c_2) boundary is either O'Brien-Fleming (OF) (in which case $c_1 = \sqrt{2}c_2$) or Pocock (PO) (in which case $c_1 = c_2$). For each choice of the primary boundary, we considered the same two choices for the secondary boundary: O'Brien-Fleming (OF) (in which case $d_1 = \sqrt{2}d, d_2 = d$) or Pocock (PO) (in which case $d_1 = d_2 = d$). In Table 1 we have tabulated the optimum values of d with the associated $1 - \varepsilon$ for selected values of r and $n = 20, 50$ and 100. For comparison purposes we have also included the corresponding values of d (taken from Tamhane et al. (2010), Table 1) for known ρ.

The Table 1 shows how the optimum confidence coefficient $1 - \varepsilon$ varies with the observed sample correlation coefficient r. We see that $1 - \varepsilon$ decreases as r increases. The explanation for this is as follows: As r increases, the upper confidence limit ρ^* gets close to 1 if $1 - \varepsilon$ becomes larger which makes the first term in (5) larger while the second term only decreases slightly because of the decrease in ε since FWER$(\Delta_1^*(1), 1)$ is fixed. Hence, to compensate for the increase in ρ^* as a result of increase in r and consequent increase in FWER$(\Delta_1^*(\rho^*), \rho^*)$, the confidence coefficient $1 - \varepsilon$ must decrease.

4. Power Comparisons

To assess the advantage of using the narrower secondary boundary resulting from the confidence limit method compared to the conservative boundary assuming the least favorable value of $\rho = 1$, we computed secondary powers for both the methods. For comparison we also computed secondary powers according to the known ρ assumption. The power computations were made for different true values of $\rho = 0.1(0.1)1.0$, $\Delta_1 = \delta_1\sqrt{n_1} = 3$, $\Delta_2 = \delta_2\sqrt{n_1} = 2$, $n_1 = n_2 = n = 20, 50, 100$, and for four different combinations of OF and PO primary and secondary boundaries. For given ρ and n_1, we generated 10,000 values of the sample correlation coefficient r from the approximate distribution of r. For each realization of r, we calculated the optimum value of d by interpolating in Table 1; the optimum (d_1, d_2) were obtained as explained before. Finally the average of the secondary powers was calculated.

The results are shown in Table 2. We have plotted the secondary powers versus the ρ values for $n_1 = n_2 = 20$ to illustrate the difference between the three methods (ρ= true ρ, $\rho = \rho^*$ and $\rho = 1$). To save space, only OF1-PO2 boundary comparisons are provided here. There are substantial power gains using the confidence limit method compared to using the most conservative approach (Figure 2).

Table 1: The d-values for the optimum secondary critical boundary d ($d_1 = \sqrt{2}d, d_2 = d$ for OF, $d_1 = d_2 = d$ for PO) and the associated confidence level for $n_1 = n_2 = n$ using the sample correlation coefficient r based on the first stage sample of size n $1 - \varepsilon$ ($\alpha = 0.05$)

Procedure	n	r								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
OF1-OF2	20	1.489	1.503	1.517	1.532	1.547	1.565	1.586	1.611	1.635
		(0.94)	(0.93)	(0.92)	(0.91)	(0.89)	(0.86)	(0.79)	(0.69)	(0.64)
	50	1.463	1.477	1.492	1.507	1.524	1.543	1.564	1.590	1.622
		(0.97)	(0.97)	(0.96)	(0.96)	(0.95)	(0.94)	(0.92)	(0.87)	(0.81)
	100	1.450	1.464	1.478	1.494	1.512	1.531	1.553	1.579	1.612
		(0.98)	(0.98)	(0.98)	(0.97)	(0.97)	(0.97)	(0.96)	(0.96)	(0.96)
	∞ $(r = \rho)$[‡]	1.416	1.428	1.440	1.455	1.473	1.493	1.519	1.551	1.591
OF1-PO2	20	1.713	1.724	1.735	1.746	1.758	1.771	1.786	1.805	1.832
		(0.95)	(0.94)	(0.93)	(0.93)	(0.92)	(0.91)	(0.89)	(0.81)	(0.65)
	50	1.692	1.703	1.715	1.727	1.740	1.755	1.771	1.791	1.822
		(0.97)	(0.97)	(0.96)	(0.96)	(0.96)	(0.95)	(0.94)	(0.91)	(0.83)
	100	1.681	1.692	1.704	1.716	1.730	1.745	1.762	1.783	1.811
		(0.98)	(0.98)	(0.98)	(0.98)	(0.97)	(0.97)	(0.97)	(0.96)	(0.91)
	∞ $(r = \rho)$[‡]	1.652	1.663	1.673	1.686	1.699	1.717	1.735	1.760	1.791
PO1-OF2	20	1.368	1.383	1.397	1.412	1.429	1.447	1.465	1.490	1.516
		(0.94)	(0.92)	(0.91)	(0.90)	(0.87)	(0.84)	(0.83)	(0.76)	(0.71)
	50	1.341	1.355	1.370	1.387	1.407	1.426	1.447	1.471	1.505
		(0.97)	(0.97)	(0.96)	(0.94)	(0.91)	(0.89)	(0.88)	(0.87)	(0.79)
	100	1.327	1.341	1.356	1.372	1.391	1.410	1.434	1.459	1.494
		(0.98)	(0.98)	(0.97)	(0.97)	(0.96)	(0.96)	(0.94)	(0.93)	(0.91)
	∞ $(r = \rho)$[‡]	1.290	1.304	1.317	1.333	1.350	1.372	1.396	1.429	1.470
PO1-PO2	20	1.697	1.707	1.717	1.729	1.741	1.756	1.774	1.793	1.818
		(0.96)	(0.95)	(0.95)	(0.93)	(0.91)	(0.89)	(0.85)	(0.84)	(0.80)
	50	1.678	1.687	1.697	1.709	1.722	1.739	1.759	1.781	1.809
		(0.98)	(0.98)	(0.97)	(0.96)	(0.96)	(0.92)	(0.88)	(0.85)	(0.82)
	100	1.669	1.677	1.687	1.699	1.712	1.727	1.745	1.768	1.802
		(0.99)	(0.99)	(0.98)	(0.98)	(0.97)	(0.96)	(0.95)	(0.94)	(0.89)
	∞ $(r = \rho)$[‡]	1.648	1.655	1.661	1.672	1.683	1.698	1.716	1.742	1.777

[†] The parenthetical entry below each optimum d is the corresponding confidence coefficient $1 - \varepsilon$.

Table 2: Secondary power comparison between the confidence limit method, the conservative method using $\rho = 1$ and the known true ρ method.

Procedure	ρ	ρ									
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
OFOF	$\rho = \rho$	0.6162	0.6204	0.6244	0.6269	0.6275	0.6266	0.6212	0.6099	0.5877	0.5254
	$\rho = \rho^*$	0.5853	0.5869	0.5883	0.5898	0.5900	0.5886	0.5847	0.5762	0.5604	0.5254
	$\rho = 1$	0.4958	0.5029	0.5097	0.5161	0.5220	0.5273	0.5315	0.5338	0.5322	0.5254
OFPO	$\rho = \rho$	0.7045	0.7091	0.7142	0.7186	0.7234	0.7268	0.7308	0.7327	0.7324	0.6954
	$\rho = \rho^*$	0.6838	0.6885	0.6929	0.6977	0.7023	0.7071	0.7112	0.7145	0.7144	0.6945
	$\rho = 1$	0.6270	0.6344	0.6420	0.6497	0.6576	0.6659	0.6746	0.6837	0.6931	0.6945
POOF	$\rho = \rho$	0.6173	0.6171	0.6172	0.6156	0.6129	0.6067	0.5977	0.5807	0.5540	0.4861
	$\rho = \rho^*$	0.5830	0.5797	0.5773	0.5746	0.5712	0.5662	0.5585	0.5459	0.5261	0.4861
	$\rho = 1$	0.4747	0.4801	0.4850	0.4894	0.4931	0.4958	0.4971	0.4963	0.4924	0.4861
POPO	$\rho = \rho$	0.6680	0.6712	0.6748	0.6766	0.6783	0.6782	0.6762	0.6691	0.6538	0.6020
	$\rho = \rho^*$	0.6544	0.6545	0.6551	0.6559	0.6567	0.6562	0.6543	0.6484	0.6349	0.6020
	$\rho = 1$	0.5846	0.5896	0.5945	0.5992	0.6036	0.6075	0.6104	0.6115	0.6091	0.6020

† For each OF-PO combination, the table entries are the secondary powers corresponding to the (d_1, d_2) boundary computed under three cases. Top row: use the known true ρ; middle row: use the confidence limit method; bottom row: use the least favorable value $\rho = 1$.

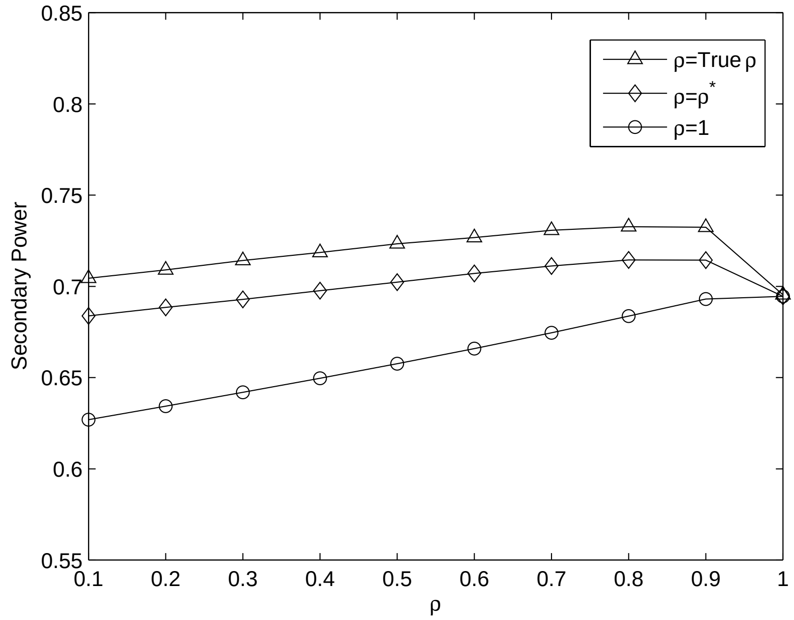

Figure 2: Secondary powers of the three methods as functions of ρ using the OF1-PO2 boundary for $n_1 = n_2 = 20$

5. Example

To illustrate the proposed methodology, we use an example from Pocock, Geller and Tsiatis (1987). A randomized double-blind crossover trial was conducted with 17 patients having asthma or chronic obstructive airways disease to compare an inhaled active drug versus placebo. Each patient received the active drug and placebo in a random order for consecutive 4-week periods. Three efficacy endpoints were measured at the end of each 4-week period. They were forced vital capacity (FVC), forced expiratory volume (FEV_1) and peak expiratory flow rate (PEFR). We will assume that FVC is the primary endpoint, FEV_1 is the secondary endpoint and ignore PEFR. Period effect was not significant, so simple matched pair t-statistics were computed on the mean differences between the drug and placebo.

	FVC	FEV_1
Mean Difference	4.81%	7.56%
Standard Deviation	10.84%	18.53%
Test Statistic	1.83	1.68

Furthermore the correlation coefficient between FEV_1 and FVC was 0.095.

Assuming that these data were obtained at the interim look of a two-stage group sequential trial with equal sample sizes $n_1 = n_2 = 17$ and using one-sided OF1-PO2 boundaries with $\alpha = 0.05$, we have $(c_1, c_2) = (1.678\sqrt{2}, 1.678)$. Since the primary test statistic $X_1 = 1.77 < c_1$, we do not reject the null hypothesis H_1 about the efficacy of the drug for FVC. The secondary boundary calculated using the confidence limit method equals $(d_1, d_2) = (1.712, 1.712)$. The corresponding confidence coefficient for ρ was $1-\varepsilon = 0.95$ and the upper confidence limit was $\rho^* = 0.551$. Check that this secondary boundary and the confidence coefficient $1 - \varepsilon$ are nearly equal to those listed in Table 1 for the OF1-PO2 boundary for $n = 20$ and $r = 0.1$. The non-adaptive (a priori) secondary boundary is

$(d_1, d_2) = (1.876, 1.876)$. Thus, suppose that at the end of the second stage the statistics were the same as those at the end of the first stage, i.e., $(X_2, Y_2) = (1.83, 1.68)$. Then $X_2 > c_2 = 1.678$ and so H_1 will be rejected but $Y_2 < 1.712$, so H_2 will not be rejected.

6. Concluding Remarks

Tamhane et al. (2010) assumed that the correlation coefficient ρ between the primary and the secondary endpoint is either an unknown nuisance parameter or a known constant. Under the former assumption they showed that $\rho = 1$ is the least favorable value of ρ and used it to compute the critical boundaries of the GSP. They also computed the critical boundaries for selected known values of ρ. However, neither assumption is very practical. Therefore we propose to use the sample correlation coefficient r from the first stage data in place of ρ to adaptively adjust the secondary boundary. The sampling error is taken into account via an upper confidence limit on ρ. One possible future extension is that since this procedure is an 'unknown correlation coefficient adjusted' GSP, which is guaranteed at level α, combining it with other adaptive GSPs such as Tamhane, Wu and Mehta (in manuscript) will result in a new GSP which will provide Type I error protection against unknown correlation coefficient.

REFERENCES

Cui, L., Hung, H. M. and Wang, S. J. (1999). Modification of sample size in group sequential clinical trials. *Biometrics*, **55**, 853-857.

Hung H. M. J., Wang, S-J and O'Neill, R. (2007). Statistical considerations for testing multiple endpoints in group sequential or adaptive clinical trials. *Journal of Biopharmaceutical Statistics*, **17**, 1201-1210.

Jennison, C. and Turnbull, B.W. (2000). Group Sequential Methods with Applications to Clinical Trials. FL: Chapman and Hall/CRC Press.

Lan, K. K. G. and DeMets, D. L. (1983). Discrete sequential boundaries for clinical trials. *Biometrika* **70**, 659-663.

O'Brien, P. C. and Fleming, T. R. (1979). A multiple testing procedure for clinical trials. *Biometrics*, **35**, 549-556.

Pocock, S. J. (1977). Group sequential methods in the design and analysis of clinical trials. *Biometrika*, **64**, 191–199.

Pocock, S. J., Geller, N. L. and Tsiatis, A. A. (1987). The analysis of multiple endpoints in clinical trials. *Biometrics*, **43**, 487–498.

Tamhane, A. C., Mehta, C. R. and Liu, L. (2010). Testing a primary and secondary endpoint in a group sequential design. *Biometrics*, **4**, 1174–1184.

Tamhane, A. C., Wu, Y. and Mehta, C. R. (in manuscript). Adaptive Extensions of the Two-Stage Group Sequential Procedure for Testing a Primary and a Secondary Endpoint (II): Sample Size Adjustment.