SAMPLE SIZE RE-ESTIMATION FOR ADAPTIVE SEQUENTIAL DESIGN IN CLINICAL TRIALS

Ping Gao1, James H. Ware2, and Cyrus Mehta2,3

1 Biostatistics, The Medicines Company, Parsippany, New Jersey, USA
2 Harvard School of Public Health, Boston, Massachusetts, USA
3 Cytel Corporation, Cambridge, Massachusetts, USA

There is considerable interest in methods that use accumulated data to modify trial sample size. However, sample size re-estimation in group sequential designs has been controversial. We describe a method for sample size re-estimation at the penultimate stage of a group sequential design that achieves specified power against an alternative hypothesis corresponding to the current point estimate of the treatment effect.

Key Words: Adaptive design; Brownian motion; Conditional power; Markov process; Sample size re-estimation; Sequential design; Transition function.

INTRODUCTION

When designing clinical trials, investigators often estimate the sample size from limited information about the variance of the response and the size of the treatment effect. The reliability of sample size estimation can be limited by several factors: the prior information may have been derived from small trials that do not provide robust estimates of the treatment effect and its variance; changes in medical practice may alter the treatment effect; the patient population in the current trial may differ from that in earlier trials; or the effect size estimate may have been derived from clinical trials of other drugs in the same class as the test drug. In such circumstances, it would be desirable to have the ability to modify the sample size during the course of the study to achieve desired power against an alternative hypothesis derived from the accumulating data. Different approaches have been proposed in the literature (e.g., Cui et al., 1999; Fisher, 1998; Jennison and Turnbull, 2003; Liu et al., 2008; Proschan and Hunsberger, 1995). Müller and Schäfer describe a general method for modifying sequential designs based on the principle of preservation of conditional rejection probabilities (Müller and Schäfer, 2001, 2004). Their approach is both general and flexible. However, in the typical situation, it requires further specification of the revised sequential design and numerical integration to determine the stopping boundaries.
In some special situations, including two-stage adaptive designs (Fisher, 1998) and sequential analysis (Cui et al., 1999), methods for sample size re-estimation with closed form solutions have been described. These methods have, however, been controversial. In particular, the method proposed by Cui et al. (1999) has been understood to require “down-weighting” of the data acquired after the sample size re-estimation, thus raising concerns about both the appropriateness and the efficiency of the method (Fleming, 2006).

In this study, we employ the properties of Brownian motion to derive a method of sample size re-estimation at the penultimate analysis in a group-sequential design. We show that the proposed adaptive design is a conventional group-sequential design with the feature that the timing of the last analysis is determined stochastically. The method is thus similar in spirit to the methods proposed by Lan and DeMets (1983) for modifying the timing of interim analyses. The proposed method is shown to be an extension of the variance spending function proposed by Fisher (1998) to the group-sequential setting. The method preserves the conditional type I error, thus conforming to the general principle described by Müller and Schäfer (2001, 2004). We show that, in the special case with only one interim analysis, the proposed method is a two-stage adaptive design of the type described by Proschan and Hunsberger (1995) with a “linear error function.” We also show that the method is equivalent to the method proposed by Cui et al. (1999) and that both are special cases of the method of Müller and Schäfer (2001, 2004). There have been debates about whether the “down-weighting” in Cui, Hung, and Wang is ethical or efficient (e.g., Jennison and Turnbull, 2003). The equivalence of the three methods demonstrates that the sample size re-estimation method of Cui, Hung, and Wang is valid and does not truly down-weight any portion of the data.

A GROUP-SEQUENTIAL DESIGN WITH SAMPLE SIZE ADJUSTMENT

We make the usual assumptions for a group-sequential design. The test statistic constructed from the accumulating data is assumed to have the properties of a Brownian motion, $W(t)$. Under the null hypothesis, $E[W(t)] = 0$ and $\text{var}[W(t)] = t$. For a specified alternative hypothesis, $W(t)$ is assumed to have drift $\theta$, that is, $E[W(t)] = \theta$, and $\text{var}[W(t)] = t$. Further, for any $s > 0$, $t > 0$, $W^*(t) = W(t+s) - W(s)$ is normally distributed and independent of $W(s)$ (Chung, 1982). In clinical trials, appropriately chosen test statistics can frequently be approximated by a Brownian motion (Jennison and Turnbull, 2000; Moyé, 2006; Whitehead, 1997). In this paper, the variance $t$ is taken to be Fisher’s information time (Whitehead, 1997), and we do not assume that $t \leq 1$ (i.e., $t$ is not “scaled”). Typically, a sequential design tests the one-sided alternative hypothesis $\theta > 0$ against the null hypothesis $\theta = 0$. We assume that a total of $K$ tests are to be performed, with upper critical values $c_1, \ldots, c_k, \ldots, c_K$ and futility (lower) boundaries $d_1, \ldots, d_{K-1}$. The trial will be stopped for futility if $Z(t_k) = W(t_k)/\sqrt{t_k} \leq d_k$ for some $1 \leq k \leq K - 1$ and stopped for efficacy if $Z(t_k) = W(t_k)/\sqrt{t_k} \geq c_k$ for some $1 \leq k \leq K$. The type I error under the null hypothesis is then

$$P[\{Z(t_1) \geq c_1\} \cup \{d_1 < Z(t_1) < c_1, Z(t_2) \geq c_2\} \cdots \cup \{d_{K-1} < Z(t_{K-1}) < c_{K-1}, Z(t_K) \geq c_K\}]$$
(Jennison and Turnbull, 2000). This probability can be written as

\[ P(Z(t_1) \geq c_1) + P(d_1 < Z(t_1) < c_1, Z(t_2) \geq c_2) + \cdots + P(d_1 < Z(t_1) < c_1, \ldots, d_{K-1} < Z(t_{K-1}) < c_{K-1}, Z(t_K) \geq c_K) \]  

(1)

**ADJUSTING THE FINAL CRITICAL VALUE AFTER SAMPLE SIZE ADJUSTMENT**

Suppose that the sample size is modified after the \( K - 1 \) st interim analysis. Then the final analysis will be performed at a new information time \( \tau_K \) (when the sample size is \( N_{\text{new}} \) instead of the originally planned \( t_K \) (with sample size \( n \). If \( \tau_K \) is a specified function of the observed data, then \( \tau_K \) is a random variable. Müller and Schäfer (2001, 2004) show that a sequential design may be modified at any of the \( K - 1 \) pre-specified analysis times, provided that the conditions for early termination have not been met and that the original and modified design have identical Type I errors conditional on the observed data. When the modification occurs at time \( t_{K-1} \), a closed form expression can be derived for the critical value at the final analysis. When the final critical value, \( c'_K \), is a pre-specified function of the available data, it is then also a random variable. As is shown in the appendix, the condition of Müller and Schäfer is met if

\[ P(d_1 < Z(t_1) < c_1, \ldots, d_{K-1} < Z(t_{K-1}) < c_{K-1}, Z(t_K) \geq c_K) = P(d_1 < Z(t_1) < c_1, \ldots, d_{K-1} < Z(t_{K-1}) < c_{K-1}, Z(t_K) \geq c_K) \]

This can be achieved by requiring \( P(Z(t_K) \geq c_K \mid Z(t_{K-1})) = P(Z(t_K) \geq c'_K \mid Z(t_{K-1})) \), and the adjusted final critical value is then given by

\[ c'_K = \frac{1}{\sqrt{\tau_K}} \left[ \frac{\sqrt{\tau_K - t_{K-1}} (c_K \sqrt{\tau_K} - \sqrt{\tau_K - t_{K-1}} Z(t_{K-1})) + \sqrt{\tau_K - t_{K-1}} Z(t_{K-1})}{\sqrt{\tau_K - t_{K-1}}} \right] \]

(2)

where \( t_K \) is the Fisher’s information time for the originally scheduled final analysis.

The Müller and Schäfer condition can also be met at the \( L \)th interim analysis by requiring

\[ P([Z(t_{L+1}) \geq c_{L+1}] \cup [d_{K-1} < Z(t_{K-1}) < c_{K-1}, Z(t_K) \geq c_K] \mid Z(t_L) = z_L) = P([Z(t_{L+1}) \geq c_{L+1}] \cup [d_{K-1} < Z(t_{K-1}) < c_{K-1}, Z(t_K) \geq c_K] \mid Z(t_L) = z_L) \]

where \( 1 \leq L \leq K - 1 \). That is, the timing of the interim analyses is changed from information time points, \( t_{L+1}, \ldots, t_K \), to new information time points, \( t'_{L+1}, \ldots, t_K \), and the critical values are modified from \( c_{L+1}, \ldots, c_K \), \( d_{K-1}, \ldots, d_{K-1} \) to new values \( c'_{L+1}, \ldots, c'_K \) and \( d'_K \) that maintain the conditional type I error. This more general approach is valid provided that the design has not been modified before the \( L \)th interim analysis. However, unless \( L = K - 1 \), the choices for \( t'_i, c'_i \), and \( d'_i \) are not unique and numerical integration is required to determine these quantities. Hence, we consider the case \( L = K - 1 \) in the remainder of this paper.
CONDITIONAL POWER UNDER THE ALTERNATIVE HYPOTHESIS

Under the alternative hypothesis $H_a$: $\theta > 0$, $Z(t) \sim N(0, \sqrt{t}^1, 1)$, and for any $s < t$, $[Z(t)\sqrt{t} - \theta t] - [Z(s)\sqrt{s} - \theta s] \sim N(0, t - s)$. Therefore, given an observed value of $Z(t_{K-1})$, the conditional probability of rejecting the null hypothesis at information time $t_K$ is

$$P(Z(t_K) \geq z_K \mid Z(t_{K-1}) = z_{K-1}) = P\left(\frac{Z(t_K) - \theta t_K}{\sqrt{t_K - t_{K-1}}} - \frac{Z(t_{K-1}) - \theta t_{K-1}}{\sqrt{t_{K-1}}} \geq \frac{c_K\sqrt{t_K - z_{K-1}\sqrt{t_{K-1}}}}{\sqrt{t_K - t_{K-1}} - \sqrt{t_{K-1}}} \mid Z(t_{K-1}) = z_{K-1}\right)$$

$$= \Phi\left(\frac{\theta(t_K - t_{K-1}) - c_K\sqrt{t_K - z_{K-1}\sqrt{t_{K-1}}}}{\sqrt{t_K - t_{K-1}} - \sqrt{t_{K-1}}}\right)$$

(3)

Similarly, the conditional power at the modified information time $\tau_K$, with the adjusted stopping boundary $c'_K$, will be

$$\Phi\left(\frac{\theta(t_K - t_{K-1}) - c'_K\sqrt{t_K - z_{K-1}\sqrt{t_{K-1}}}}{\sqrt{t_K - t_{K-1}} - \sqrt{t_{K-1}}}\right)$$

$$= \Phi\left(\theta\sqrt{t_K - t_{K-1}} - \frac{1}{\sqrt{t_K - t_{K-1}}} (c_K\sqrt{t_K - \sqrt{t_{K-1}\sqrt{z_{K-1}}}})\right)$$

(4)

Therefore, for $\theta > 0$ the conditional power will approach 100% if the sample size can be increased without limit. If $\theta$ is estimated at the interim analysis, the estimate can be used to determine conditional power for that alternative hypothesis by substituting the estimate in the previous formula. For any desired power, $(1 - \beta) \times 100\%$, the information time, $\tau_K$, required to achieve desired power against a specified alternative, $\theta$, can be determined by solving the equation $1 - \beta = \Phi(\theta\sqrt{t_K - t_{K-1}} - \frac{1}{\sqrt{t_K - t_{K-1}}} (c_K\sqrt{t_K - \sqrt{t_{K-1}\sqrt{z_{K-1}}}}))$.

Thus, the required information time is given by

$$\tau_K = \frac{1}{\theta^2} \left(\frac{1}{\sqrt{t_K - t_{K-1}}} (c_K\sqrt{t_K - \sqrt{t_{K-1}\sqrt{z_{K-1}}}}) + Z_\beta\right)^2 + t_{K-1}$$

(5)

When the sample size is large, the information time is approximately proportional to the sample size (Whitehead, 1997). Hence the new sample size, $N_{\text{new}}$, corresponding to information time $\tau_K$, can be approximated by

$$N_{\text{new}} \approx \frac{\tau_K}{t_K} n$$

(6)

where $n$ is the original planned sample size (corresponding to information time $t_K$). The required sample size adjustment can also be expressed as

$$\gamma = \frac{t_K - t_{K-1}}{t_K - t_{K-1}} = \frac{(c_K\sqrt{t_K - \sqrt{t_{K-1}\sqrt{z_{K-1}}}} + Z_\beta\sqrt{t_K - t_{K-1}})^2}{(t_K - t_{K-1})^2 \theta^2}$$
(See Jennison and Turnbull, 2003). We note that there is no restriction on the range of \( \gamma \), whereas restrictions were required by Jennison and Turnbull.

Our formulas can be expressed in terms of scaled information time. Let \( s_i = \frac{1}{t_i} \). Then \( s_i \leq 1 \). (5) becomes

\[
\tau_K = \frac{1}{\hat{\theta}^2} \left( \frac{1}{\sqrt{1 - s_{K-1}}} (c_K - \sqrt{s_{K-1}} z_{K-1}) + Z_{\beta} \right)^2 + t_{K-1}
\]

and

\[
\gamma = \frac{\tau_K - t_{K-1}}{t_K - t_{K-1}} = \frac{(c_K - \sqrt{s_{K-1}} z_{K-1})^2}{(t_K - t_{K-1})(1 - s_{K-1})\hat{\theta}^2}
\]

\[
N \approx \frac{\tau_K}{t_{K-1}} s_{K-1} n
\]


Fisher (1998) described a conditional power-based sample size calculation in a setting that includes a final analysis and interim analyses without early termination. For the sequential design, Cui et al. (1999) proposed an unconditional sample size calculation, with the new sample size given by \( N_{new} = \left( \frac{Z}{\hat{\theta}} \right)^2 n \), where \( \hat{\theta} \) is the effect size assumed in the initial study design, \( Z \) is the estimated effect size from the observed data, and \( n \) is the initially planned sample size. Later we show in an example that the method of Cui, Hung, and Wang may not provide adequate conditional power. Jennison and Turnbull (2003) discussed sample size estimation for a sequential design that placed restrictions on the range of \( \gamma = \frac{N_{new} - n}{1 - r} \), where \( 0 < r < 1 \), and \( mn \) is the sample size at which the interim analysis is performed. We show that the sample size can be re-estimated in the sequential design setting with no restriction on the range of \( \gamma \). This method reduces to Fisher’s sample size estimation when no early termination is planned.

Let \( S_1 = W(t_{K-1}) \), and \( S_2 = W(\tau_K) - W(t_{K-1}) \). In the special case of differences between normally distributed variables \( X_A \) and \( X_B \), suppose the \( K-1 \) st interim analysis is performed when the sample size is \( mn \) with \( r < 1 \): \( S_1 = \sum_{i=1}^{mn} (X_{Ai} - X_{Bi}) \), \( S_2 = \sum_{i=mn+1}^{N_{new}} (X_{Ai} - X_{Bi}) \), and \( \gamma = \frac{N_{new} - mn}{(1 - r)mn} \) (Jennison and Turnbull, 2003). We have (after some basic algebraic manipulations)

\[
P(Z(\tau_K) \geq c_K \mid Z(t_{K-1}) = z_{K-1})
\]

\[
= P\left( Z(\tau_K) \geq \frac{1}{\sqrt{\tau_K}} \left[ \frac{\sqrt{\tau_K} - t_{K-1}}{\sqrt{t_K - t_{K-1}}} (c_K \sqrt{t_K} - \sqrt{t_{K-1}}z_{K-1}) + \sqrt{t_{K-1}}z_{K-1} \right] \mid Z(t_{K-1}) = z_{K-1} \right)
\]

\[
= P\left( \frac{\gamma - \frac{1}{\sqrt{t_K}} S_2 + S_1}{\sqrt{t_K}} \geq c_K \mid Z(t_{K-1}) = z_{K-1} \right)
\]

(9)
In the special case of normally distributed variables with variance $\sigma^2 = 1$ (Cui et al., 1999; Jennison and Turnbull, 2003), $t_K = n$, $t_{K-1} = m$, and $\tau_K = N_{\text{new}}$. Equation (9) becomes $\frac{\hat{\theta} + \sqrt{\hat{\theta}^2 + 4\gamma^2}}{2} \geq \gamma$. This is identical to the adjustment made in Cui et al. (1999) and in Fisher (1998) when there is no early termination. Therefore, adjusting the final critical value to $c'_K$ is equivalent to the adjustment of Cui, Hung, and Wang when $L = K - 1$. In the case of no early termination, this method of adjustment is also equivalent to Fisher’s variance spending design (Jennison and Turnbull, 2003).

The availability of an explicit formula for $c'_K$ makes it possible to modify the sample size to achieve desired conditional power against a specified alternative hypothesis, whereas the adjusted sample size suggested by Cui, Hung, and Wang does not, as illustrated in our example. In the special case of no early stopping, $c_K = \Phi^{-1}(1 - \alpha)$. If the comparison is between normally distributed variables with variance $\sigma^2 = 1$ and $\hat{\theta} = \frac{m}{n}$, then we have $\gamma = \frac{(\frac{\hat{\theta}}{m} - \gamma)(\frac{\hat{\theta}}{m})^2}{\gamma(1 - \gamma^2)}$ (Jennison and Turnbull, 2003). Hence, our sample size re-estimation extends the method of Fisher to sequential designs.

**TIMING OF SAMPLE SIZE MODIFICATION**

The sample size modification procedure discussed here is performed at the $K - 1$st interim analysis. Thus it is not as general as the Müller and Schäfer procedure with which sample size modification can be performed at any $L$th interim analysis with $1 \leq L < K - 1$. However, this limitation may not be material in practice. Suppose that an effect size of $\theta_0$ was assumed in the original design, and at the $L$th interim analysis the estimate is $\hat{\theta}_L$. If $\hat{\theta}_L > \theta_0$, sample size reduction may not be desirable because of concerns about the robustness of the estimate. Also, sample size reduction may not be necessary because the original sequential test procedure already allows for early termination if the actual effect size is larger than $\theta_0$. If $\hat{\theta}_L < \theta_0$, a sample size increase may be needed. But the decision on sample size increase does not need to be made at this time. Instead, the required sample size is likely to increase beyond the original planned sample size, the decision can be delayed until the $K - 1$st interim analysis, when more data is available and the estimate is more robust, thus the sample size re-estimation is more reliable. Therefore, there is no loss of flexibility if sample size modification is not performed at any time point prior to the $K - 1$st interim analysis. The availability of closed formulas at the $K - 1$st interim analysis makes the procedure easier to use.

**COMPARISON WITH PROSCHAN AND HUNSBERGER (1995)**

Proschan and Hunsberger (1995) proposed a two-stage design in which a conditional error function $A(z)$ is selected such that $A(z)$ is increasing and $\int_{-\infty}^{\infty} A(z) \varphi(z) dz = \alpha$.

To compare the two approaches, we set $K = 2$. Suppose that a futility stopping boundary $d_1$ is also selected. Then the type I error of our adaptive sequential design will be

$$\int_{d_1}^{\infty} \varphi(z) dz + \int_{d_1}^{c_1} \left[ 1 - \Phi \left( \frac{c_2 \sqrt{t_2} - \sqrt{t_1} \hat{\theta}}{\sqrt{t_2 - t_1}} \right) \right] \varphi(z) dz = \alpha$$

(10)
Therefore, an adaptive sequential design with stopping for futility is equivalent to a two-stage adaptive design with the conditional error function

\[
A(z_1) = \begin{cases} 
1 & \text{if } z_1 \geq c_1 \\
1 - \Phi\left(\frac{c_2\sqrt{t_2} - \sqrt{t_2 - t_1} z_1}{\sqrt{t_2 - t_1}}\right) & \text{if } d_1 < z_1 < c_1 \\
0 & \text{if } z_1 \leq d_1 
\end{cases}
\]

Hence, the conditional error function is of the “linear” type, with \(a = \frac{c_2\sqrt{t_2}}{\sqrt{t_2 - t_1}}\) and \(b = -\frac{\sqrt{t_1}}{\sqrt{t_2 - t_1}}\). For example, if the interim analysis is planned when 75% of the information is available, then \(t_1 = 0.75t_2\), \(a = 2c_2\), and \(b = -\sqrt{3}\), where \(c_2\) is the critical value at \(t_2\). Thus the two-stage adaptive design with an appropriate linear error function is an adaptive sequential design with only one interim analysis.

**EXAMPLE**

Suppose that a trial is being conducted to demonstrate that a new treatment is superior to a control treatment. The end point is success, defined as being free of any of a specified set of cardiac events. For binary data, the “drift” parameter is the log-odds ratio (Whitehead, 1997). Suppose that the investigators assumed initially that the success rate in the control arm would be 90%, and the success rate in the test arm would be 93.5%. Suppose also that a group sequential analysis is planned, with one interim analysis when 75% of the data are available. O’Brien and Fleming (1979) critical values are to be used and a sample size re-estimation will be performed at the interim analysis. The (traditional) sequential analysis scheme is shown in Table 1.

In this design, \(c_1 = 2.340\) and \(c_2 = 2.012\). Suppose that when the first interim analysis is performed with 783 patients in each arm, the observed cardiac event rate is only 7.3% (success rate 92.7%, number of successes = 726) in the control arm and 5.75% (success rate 94.25%, number of successes = 738) in the test arm. Using a formula from Whitehead (1997), Fisher’s information time is \(t_1 = t_{K-1} = 23.839\). The value of the test statistic, \(Z(t_1)\), is \(1.229 < c_1\). The estimated log odds ratio is given by \(\hat{\theta} = 0.253\).

Suppose now that the sample size is re-estimated. This can be done in one of two ways:

- Use Equations (5)–(6). Substitute \(c_2 = 2.012\), \(t_1 = t_{K-1} = 23.839\), and (assuming the current trend continues) \(t_K = t_2 = \frac{t_1}{0.75}\), \(Z(t_1) = z_1 = 1.229\), \(\beta = 0.2\), and

**Table 1** Sequential design scheme

<table>
<thead>
<tr>
<th>Look</th>
<th>Time (scaled)</th>
<th>Upper boundary</th>
<th>Nominal (\alpha) spending</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.75</td>
<td>2.340</td>
<td>0.0096</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2.012</td>
<td>0.0221</td>
</tr>
</tbody>
</table>

Spending rule = O’Brien–Fleming, \(P_1 = 0.1\), \(P_2 = 0.065\), \(N_1 = N_2 = 1044\), Power = 80%, \(\alpha = 0.025\) one-sided.
\[
\theta = 0.253. \text{ We calculate that the scaled information time, } \tau_\theta = 4.437 \times t_2, \text{ equal to } 4,633 \text{ patients per arm, will provide } 80\% \text{ power. With the sample size modification, the adjusted final critical value is } c'_2 = 2.236 > c_2. \\
\]

- Use the estimate suggested by Cui et al. (1999). The original assumed effect size (log-odds ratio) is \( \theta = 0.469 \). The estimated effect size is \( \hat{\theta} = 0.253 \). Hence, the new sample size should be \( (\theta/\hat{\theta})^2 = 3.44 \) times the original sample size, or 2,694 patients per arm. Equation (4) shows that this sample size provides a power of 67.1%.

**A SIMULATION EXPERIMENT**

In this section we demonstrate through a massive simulation experiment that our proposed method of preserving type-1 error after an unblinded sample size re-estimation based on adjusting the final critical region (ACR method) is equivalent to the method of Cui, Hung, and Wang (CHW method) and the method of Müller and Schäfer (MS method). To this end we simulated a two-arm randomized clinical trial with normally distributed observations, having 90% power to detect a difference of means \( \theta = 0.3 \) using a one-sided level -0.05 group sequential test consisting of three equally spaced looks, with stopping boundaries derived from the one-sided Lan and DeMets (1983) error spending function:

\[
\alpha(t_k) = 2 - 2\Phi\left(\frac{z_\alpha/2}{\sqrt{n_k/n}}\right), \quad k = 1, 2, 3
\]

where the \( n_k \)'s are the planned sample sizes at the three looks. We assume a known standard deviation \( \sigma = 1 \). This trial attains the desired 90% power with planned sample sizes \( n_1 = 129, n_2 = 258, \) and \( n_3 = 388 \). Suppose it is decided at look 2 to increase the final sample size from \( n_3 \) to \( N_\text{new} \) based on the following two rules:

**Rule (1)** If \( \hat{\theta} < 0 \text{ or } \hat{\theta} > 0.3 \), the final sample size remains unchanged at \( n_3 = 390 \).

**Rule (2)** If \( 0 \leq \hat{\theta} \leq 0.3 \), determine the value \( N^* \) such that the conditional power at sample size \( N^* \) is 90%. Then the new final sample size is \( N_\text{new} = \min(1000, N^*) \).

Table 2 displays the unconditional power and expected sample size of trials in which sample size is changed according to these two rules and the type-1 error is controlled by the ACR, CHW, and MS methods. Each power and expected sample size estimate is obtained as the average of 1,000,000 simulated clinical trials. The values of \( \theta \) for which the simulations are performed range between 0 and 0.3, in step sizes of 0.05. The operating characteristics of the three methods are identical up to Monte Carlo accuracy. (The small differences observed between the three methods are attributable to using different starting seeds for the random number generation.)

**BEHAVIOR OF \( C'_k \) AS A FUNCTION OF STATISTICS AT LOOK \( K - 1 \)**

It is interesting to study the relationship between the final critical value, \( c'_k \), and some of the statistics obtained at look \( K - 1 \). Consider again the 3-look group
Table 2: Power and expected sample size (ASN) for the ACR, CHW, and MS methods

<table>
<thead>
<tr>
<th>$\theta$</th>
<th>Pwr-ACR</th>
<th>Pwr-CHW</th>
<th>Pwr-MS</th>
<th>ASN-ACR</th>
<th>ASN-CHW</th>
<th>ASN-MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.3</td>
<td>0.9831</td>
<td>0.9833</td>
<td>0.9834</td>
<td>423.11</td>
<td>423.03</td>
<td>423.35</td>
</tr>
<tr>
<td>0.25</td>
<td>0.9453</td>
<td>0.9460</td>
<td>0.9458</td>
<td>518.84</td>
<td>518.59</td>
<td>519.03</td>
</tr>
<tr>
<td>0.2</td>
<td>0.8417</td>
<td>0.8422</td>
<td>0.8416</td>
<td>615.05</td>
<td>615.38</td>
<td>615.38</td>
</tr>
<tr>
<td>0.15</td>
<td>0.6360</td>
<td>0.6362</td>
<td>0.6363</td>
<td>690.04</td>
<td>690.56</td>
<td>689.86</td>
</tr>
<tr>
<td>0.1</td>
<td>0.3702</td>
<td>0.3703</td>
<td>0.3715</td>
<td>724.67</td>
<td>724.2</td>
<td>724.8</td>
</tr>
<tr>
<td>0.05</td>
<td>0.1581</td>
<td>0.1585</td>
<td>0.1582</td>
<td>709.62</td>
<td>710.17</td>
<td>710.45</td>
</tr>
<tr>
<td>0.0</td>
<td>0.0505</td>
<td>0.0498</td>
<td>0.0497</td>
<td>655.74</td>
<td>655.92</td>
<td>656.09</td>
</tr>
</tbody>
</table>

ACR = Adjusting the Final Critical Region method; CHW = Cui, Hung, and Wang method; MS = Müller and Schäfer method.

sequential design presented previously. Suppose, however, that the final sample size $N_{\text{new}}$ at look 3 is determined entirely by Equation (5) so as to obtain 90% conditional power, i.e., the restrictions of Rule (1) and Rule (2) that were imposed for the simulation results in Table 2 are lifted. Figure 1 plots the values of $c'_k$ and $c_k$ as functions of the conditional power at look $K - 1$.

For values of conditional power between a lower and upper threshold, the adjusted critical cut-off value $c'_k$ is actually lower than the unadjusted classical group sequential one-sided level $-0.05$ cut-off $c_k = 1.6948$. A similar result can be observed in Fig. 2 by plotting $c'_k$ and $c_k$ as functions of the observed value of $\hat{\theta}_2$ at look 2 because larger values of conditional power are directly related to larger values of $\hat{\theta}_2$.

**Upper Threshold**

The upper threshold at which $c'_k$ and $c_k$ are equal occurs when the observed value of $\hat{\theta}_2$ at look 2 is good enough for the conditional power to be equal to 90% exactly with no sample size change, i.e., with $N_{\text{new}} = 388$. For the current example,
this occurs at $\hat{\theta}_2 = 0.2472$. When $\hat{\theta}_2 > 0.2472$, it is actually possible to attain the desired 90% conditional power by decreasing the final sample size $N_{new}$ to less than the initially planned 388 subjects as shown in Fig. 3.

Evidently, if the results at look 2 are so promising that one can attain the desired conditional power of 90% by decreasing the final sample size, the ACR procedure compensates by making $c'_K$ exceed $c_K$ to preserve the type 1 error. In fact one can show analytically that $c'_K \to z_{K-1}$ as $N_{new} \to n_{K-1}$.

**Lower Threshold**

Figure 2 shows that, as the observed $\hat{\theta}_2$ decreases to less than 0.2472, the final critical value $c'_K$ continues to decrease until it reaches a nadir and then begins to increase until, at $\hat{\theta}_2 = 0.1369$, the two critical values, $c'_K$ and $c_K$, are again equal. Then, as $\hat{\theta}_2$ decreases even further, $c'_K$ once again exceeds $c_K$. The corresponding

---

**Figure 2** ACR and classical critical values as a function of $\theta_{hat,2}$.

**Figure 3** New sample size needed to attain 90% conditional power for observed $\theta_{hat,2}$.
lower and upper thresholds on the conditional power scale are seen from Fig. 1 to be 27.4% and 90%, respectively.

**Comparison to Results of Chen et al. (2004)**

We have seen that, for conditional power values between 27.4% and 90%, $c'_K < c_K$. More generally, we have provided analytical expressions for computing lower and upper thresholds within which $c'_K < c_K$. It follows that, if one wishes to increase the sample size at look $K - 1$ but nevertheless perform the final hypothesis test at look $K$ with the classical group sequential boundary, $c_K$, the type-1 error will be preserved conservatively if the conditional power falls between the lower and upper thresholds. These results support and further generalize the assertion of Chen et al. (2004) that one can increase the sample size at look $K - 1$ without making any type of adaptive adjustment, provided the conditional power exceeds 50%.

**DISCUSSION**

Though there has been considerable methodological work on adaptive sample size estimation for group-sequential designs, investigators have been reluctant to employ these methods in Phase 3 studies of new therapies. The methods discussed in this paper provide a relatively straightforward method for modifying a group sequential design at the penultimate analysis to achieve the desired power against any new alternative hypothesis specified by the investigators, including alternatives determined from analysis of the accumulated data. The modified designs belong to the general family of group sequential designs for Brownian motion and they do not involve differential weighting of data accumulated in different time periods.

We have shown in Fig. 1 that if the conditional power at the penultimate look lies in a fairly wide range, the final critical boundary, $c'_K$, for the ACR test is lower than the final critical boundary, $c_K$, for the classical group sequential test. This fact has two important implications for adaptive designs: (1) Within a pre-determined range of conditional power values, we are free to increase the sample size by any amount according to our preference and budget, and we may nevertheless perform the usual classical group sequential test, $Z(t_K) \geq c_K$, at the end of the trial rather than the corresponding adaptive ACR test $Z(t_K) \geq c'_K$. This will not inflate the type-1 error because in this range of conditional power,

$$P\left\{\bigcup_{k=1}^{K-1} Z(t_k) \geq c_k \cup Z(t_K) \geq c_K\right\} \leq P\left\{\bigcup_{k=1}^{K-1} Z(t_k) \geq c_k \cup Z(t_K) \geq c'_K\right\} = \alpha$$

Furthermore, if the classical group sequential test is adopted, we may use established methods (e.g., Tsiatis et al., 1984) to compute $p$-values and confidence intervals following a group sequential test. (2) Alternatively we may perform the ACR adaptive test, $Z(t_K) \geq c'_K$. This test will always preserve the type-1 error while providing greater overall power when the conditional power at the penultimate look lies in a pre-determined range.

Finally, although this work will hopefully resolve some technical criticism of these methods, many questions remain about how sponsors, investigators, data
safety monitoring boards, and the Food and Drug Administration can work together to employ adaptive designs that meet other requirements, such as those involving blinding of study data and independence of investigators.

APPENDIX

Calculation of the Adjusted Final Critical Value

Using (1), we need

\[
P(d_1 < Z(t_1) < c_1, \ldots, d_{K-1} < Z(t_{K-1}) < c_{K-1}, Z(t_K) \geq c_K)
\]

\[
= P(d_1 < Z(t_1) < c_1, \ldots, d_{K-1} < Z(t_{K-1}) < c_{K-1}, Z(t_K) \geq c_K)
\]

or

\[
\int_{d_{K-1}}^{c_{K-1}} P(Z(t_K) \geq c_K | Z(t_{K-1}) = z_{K-1})P(d_1 < Z(t_1) < c_1, \ldots, Z(t_{K-1}) = z_{K-1})\, dz_{K-1}
\]

\[
= \int_{d_{K-1}}^{c_{K-1}} P(Z(t_K) \geq c_K | Z(t_{K-1}) = z_{K-1})P(d_1 < Z(t_1) < c_1, \ldots, Z(t_{K-1}) = z_{K-1})\, dz_{K-1}
\]

(A.1)

Equation (A.1) can be satisfied if the integrands are equal, i.e.,

\[
P(Z(t_K) \geq c_K | Z(t_{K-1}) = z_{K-1}) = P(Z(t_K) \geq c_K | Z(t_{K-1}) = z_{K-1})
\]

for all \( z_{K-1} \in (-\infty, +\infty) \) i.e., \( P(Z(t_K) \geq c_K | Z(t_{K-1})) = P(Z(t_K) \geq c_K | Z(t_{K-1})) \)

Since \( Z(t_{K-1}) = W(t_{K-1})/\sqrt{t_{K-1}} \)

\[
P(W(t_K)/\sqrt{t_K} \geq c_K | W(t_{K-1}))
\]

\[
= P(W(t_K) - W(t_{K-1}) \geq c_K \sqrt{t_K} - W(t_{K-1}) | W(t_{K-1}))
\]

\[
= P\left( \frac{W(t_K) - W(t_{K-1})}{\sqrt{t_K - t_{K-1}}} \geq \frac{c_K \sqrt{t_K} - W(t_{K-1})}{\sqrt{t_K - t_{K-1}}} \middle| W(t_{K-1}) \right)
\]

\[
= 1 - \Phi\left( \frac{c_K \sqrt{t_K} - \sqrt{t_K - 1}Z(t_{K-1})}{\sqrt{t_K - t_{K-1}}} \right)
\]

where \( \Phi(x) \) is the cumulative distribution function of the standard normal distribution. Similarly, \( P(W(t_K)/\sqrt{t_K} \geq c_K' | W(t_{K-1})) = 1 - \Phi\left( \frac{c_K' \sqrt{t_K} - \sqrt{t_K - 1}Z(t_{K-1})}{\sqrt{t_K - t_{K-1}}} \right) \).

Here, we utilized the Markov property of the Brownian motion, namely, that, given the “present” (i.e., \( W(t_{K-1}) \)), the “future” (any event after \( t_{K-1} \), including \( W(t_K) \) and \( W(t_K) \)) is independent of the “past” (i.e., \( W(t_1), \ldots, W(t_{K-2}) \)). Moreover, \( W(t_K) - W(t_{K-1}) \) and \( W(t_K) - W(t_{K-1}) \) are independent of any event before and up to \( W(t_{K-1}) \) (see Chung, 1982, Chaps. 1 and 4).

Hence, (A.1) can be satisfied if

\[
\frac{c_K \sqrt{t_K} - \sqrt{t_K - 1}Z(t_{K-1})}{\sqrt{t_K - t_{K-1}}} = \frac{c_K' \sqrt{t_K} - \sqrt{t_K - 1}Z(t_{K-1})}{\sqrt{t_K - t_{K-1}}}
\]

(A.2)
That is, if the adjusted critical value is

$$c'_K = \frac{1}{\sqrt{t_K}} \left[ \frac{\sqrt{t_K - t_{K-1}}}{\sqrt{t_K - t_{K-1}}} (c_K \sqrt{t_K} - \sqrt{t_{K-1}} Z(t_{K-1})) + \sqrt{t_{K-1}} Z(t_{K-1}) \right]$$  \hspace{1cm} (A.3)

Alternatively, (A.2) can be obtained by utilizing the transition function (see Chung, 1982, pp. 6–11)

$$p_{x,t}(x, y) = \frac{1}{\sqrt{2\pi(t-s)}} \exp \left\{ -\frac{(y-x)^2}{2(t-s)} \right\}$$

ACKNOWLEDGMENT

The authors had helpful discussions with Tommy Fu.

REFERENCES


