Sample Size for Response-Adaptive Randomization Procedures

Feifang Hu
Department of Statistics
University of Virginia
Email: fh6e@virginia.edu

Summary

Response-adaptive randomization procedures have often been proposed as a way of using accruing data to affect future allocation in clinical trials. It is difficult to calculate the requisite sample size for adaptive designs, because the allocation probabilities keep changing during the clinical trials. In this article, we focus on the sample size of two-arm (drug versus control) clinical trials. Based on asymptotic properties, we derive an approximate power function for each fixed sample size. Then formulas of sample size are derived for response-adaptive randomization procedures. These formulas are also useful for general randomization procedures. Some simulation studies also support the approximate power function as well as the estimated sample size.

Key words. Biased coin designs; Doubly adaptive biased coin designs; Power; Randomized play-the-winner rule; Sequential designs; Variation.

---

The research was supported in part by NSF grant GA 10295 and by a grant from University of Virginia. This work was partially done when the author visited the Institute of Mathematical Science, National University of Singapore in 2002. The visit was supported by the Institute and by a grant from BMRC-NSTB of Singapore.
1 Introduction

Response-adaptive randomization, an important subdivision of experimental design, is a procedure in which the probability a treatment is assigned to each patient depends upon the results of the previous patients in the study. The goals are (i) to skew assignment probabilities to favour the treatment performing better thus far in the study; or (ii) to achieve a certain power as early as possible.

The most popularly used and studied adaptive designs are those based on urn models. Urn models were early proposed and studied by Athreya and Karlin (1968), Zelen (1969), Wei (1979), Wei and Durham (1978) and others. Some recent developments are reviewed in a discussion paper by Rosenberger (2002). The main advantages are (i) they are easy to implement in clinical trial; (ii) they shift randomization probabilities to favor the better treatment; and (iii) they have well studied properties.

Another important family of response-adaptive randomization procedures are based on biased coin design (Efron, 1971). In comparing two treatments (treatment and control), when balance (50%-50%) is desired in the allocation, Efron (1971) discussed the drawbacks of complete randomization. To overcome these drawbacks, Efron (1971) and Wei (1978) proposed biased coin designs which offer a compromise between complete randomization and perfect balance to reduce experimental bias and to increase the precision of inference about treatment difference. Smith (1984) and Wei, Smythe and Smith (1986) extended these designs to multi-treatment case when balance is desired or the desired allocation proportions are known.

In applications, the desired allocation proportions are usually unknown and depend on some unknown parameters. In these cases, Eisele (1994) and Eisele and Woodroofe (1995) introduced a doubly adaptive biased coin design for comparing two treatments. Recently Hu and Zhang (2002) proposed a family of doubly adaptive biased coin design for $K \geq 2$ treatments and obtained some important properties of the doubly adaptive biased coin design.

It is usually difficult to determine the requisite sample size for response-adaptive randomization procedures. This is because the allocation probabilities change from time to time. Given a fixed sample size, the number of patients assigned to each treatment is a random variable. Therefore the power of a randomized design with a fixed sample size is also a random variable. To study the sample size required, we need to know the distribution of allocation as well as its relation to power.

In the literature, sample sizes are calculated by ignoring the randomness of the allocation. Both Eisele and Woodroofe (1995) and Rosenberger and Lachin (2002, page 26) assumed that the allocations are fixed (not random) and predetermined. For example, the Behrens-Fisher problem in Section 3 (with $\sigma_E = 1$ and $\sigma_C = 2$), the sample size is 62 (from the formula of Rosenberger and Lachin (2002)) for complete
randomization, when the target power is 0.80. For this case, if a sample size of 62 is used for a complete randomization, with 20% chance, the power is less than 0.76. This is because the sample size 62 is calculated by ignoring the randomness of the allocation.

Because the power of a randomization procedure is a random variable, we may want to find a sample size (i) to achieve a fixed power on average; or (ii) to achieve a fixed power with certain probability (confidence). In Section 2, we first derive an approximate power function $\beta_0(n)$ for a given $n$ based on the asymptotic distributions of allocations. Here the $\beta_0(n)$ is usually a random variable for a randomization procedure. Let $\beta(n) = E\beta_0(n)$, the average power for a given $n$. For a given power $\beta$, we then derive formulas to estimate the required sample sizes $n$ satisfying (i) $\beta(n) \geq \beta$ or (ii) $\beta_0(n) \geq \beta$ with certain probability respectively.

In Section 3, we apply the formulas in Section 2 to calculate the requisite sample sizes for Behrens-Fisher problem for several different response-adaptive randomization procedures. Sample sizes of response-adaptive randomization procedures for binary response trials are calculated and compared in Section 4. Simulation studies of the approximate power function are also reported in Section 3 and 4. We discuss some future research topics and some concluding remarks in Section 5. In this paper, we focus on response-adaptive randomization procedures, but the results in this paper apply to all randomization procedures.

## 2 Sample Size Calculation

We now focus on the sample size calculation for comparing two treatments (experimental and control) in clinical trials. Let $X_1, ..., X_{n_E}$ be the responses of patients on the experimental treatment and $Y_1, ..., Y_{n_C}$ be the observations from the control treatment, where $n_E$ and $n_C$ ($n_E + n_C = n$) are the number of patients in the experimental group and control group respectively. Without loss of generality, suppose $\mu_E = EX_1$ and $\mu_C = EY_1$ are the two parameters of interest. In this paper, we only consider a one-sided hypothesis testing:

$$H_0 : \mu_E = \mu_C, \quad \text{versus} \quad H_1 : \mu_E > \mu_C.$$  

Generalization to a two-sided test is straightforward.

Suppose $\hat{\mu}_E$ and $\hat{\mu}_C$ are the corresponding estimators of $\mu_E$ and $\mu_C$ based on the data. Let $\hat{\sigma}_E^2$ and $\hat{\sigma}_C^2$ denote the corresponding variance estimates of $\sqrt{n} \hat{\mu}_E$ and $\sqrt{n} \hat{\mu}_C$ respectively.

**Assumption 2.1** Assume that as $n_E \to \infty$ and $n_C \to \infty$, (i) $\hat{\sigma}_E^2 \to \sigma_E^2$ and $\hat{\sigma}_C^2 \to \sigma_C^2$
in probability; and (ii) \( \sqrt{n_E}(\hat{\mu}_E - \mu_E) \rightarrow N(0, \sigma_E^2) \) and \( \sqrt{n_C}(\hat{\mu}_C - \mu_C) \rightarrow N(0, \sigma_C^2) \) in distribution, where \( \sigma_E^2 \) and \( \sigma_C^2 \) are some positive constants.

**Remark 2.1** When the estimators \( \hat{\mu}_E \) and \( \hat{\mu}_C \) are maximum likelihood estimators, moment estimators or estimators from some estimating functions, Assumption 2.1 is usually satisfied.

Based on Assumption 1, for a given significant level \( \alpha \), we reject \( H_0 \) if

\[
\frac{\hat{\mu}_E - \hat{\mu}_C}{\sqrt{\sigma_E^2/n_E + \sigma_C^2/n_C}} > z(\alpha),
\]

where \( z(\alpha) \) is defined as \( P(Z > z(\alpha)) = \alpha \). Here \( Z \) is a standard normal random variable.

**Assumption 2.2** For a given adaptive design, we usually have the following asymptotic results:

\[
n_E/n \rightarrow \nu \text{ almost surely with } 0 < \nu < 1;
\]

and

\[
\sqrt{n}(n_E/n - \nu) \rightarrow N(0, \tau^2)
\]

in distribution for some \( \tau^2 > 0 \).

**Remark 2.2** Assumption 2.2 is usually true for randomization procedures. For example, it is satisfied for Efron’s biased coin design (Efron, 1971), adaptive biased coin designs (Wei, 1978), play-the-winner rule (Zelen, 1969) and randomized play-the-winner rule (Wei and Durham, 1978). Some other examples will be discussed in the next three sections.

To achieve the power \( \beta \), it is required that

\[
P \left( \frac{\hat{\mu}_E - \hat{\mu}_C}{\sqrt{\sigma_E^2/n_E + \sigma_C^2/n_C}} > z(\alpha) \mid H_1 \right) > \beta.
\]

That is

\[
P \left( \frac{\hat{\mu}_E - \hat{\mu}_C - (\mu_E - \mu_C)}{\sqrt{\sigma_E^2/n_E + \sigma_C^2/n_C}} > z(\alpha) - \frac{\mu_E - \mu_C}{\sqrt{\sigma_E^2/n_E + \sigma_C^2/n_C}} \mid H_1 \right) > \beta.
\]

Thus, we have

\[
z(\alpha) - \frac{\mu_E - \mu_C}{\sqrt{\sigma_E^2/n_E + \sigma_C^2/n_C}} < -z(1 - \beta).
\]
Hence
\[ \frac{\mu_E - \mu_C}{\sqrt{\sigma_E^2/n_E + \sigma_C^2/n_C}} > z_{(\alpha)} + z_{(1-\beta)}. \]

Therefore, the \( n_E \) and \( n_C \) must satisfy
\[ \frac{\sigma_E^2/n_E + \sigma_C^2/n_C}{(\mu_E - \mu_C)^2} \leq \left[ \frac{\mu_E - \mu_C}{z_{(\alpha)} + z_{(1-\beta)}} \right]^2. \]

In Rosenberger and Lachin (2002, page 26), assume that \( n_E = n\nu \) and \( n_C = n(1 - \nu) \) are predetermined. The sample size \( n_0 \) can then be calculated as
\[ n = \frac{(\sigma_E^2/\nu + \sigma_C^2/(1 - \nu))(z_{(\alpha)} + z_{(1-\beta)})^2}{(\mu_E - \mu_C)^2}. \quad (5) \]

**Remark 2.3** The sample size formula in (5) is also used in Eisele and Woodroofe (1995) for their doubly adaptive biased coin design. In their paper, the randomness of \( n_E \) is ignored. As pointed out in Rosenberger and Lachin (2002, page 26), this is not the case under response-adaptive randomization. But they did not study this problem further.

In randomization procedures, both \( n_E \) and \( n_C \) are random variables for a fixed \( n \). Based on Assumption 2.2, we have \( n_E \to \infty \) and \( n_C \to \infty \) almost surely as \( n \to \infty \). Now from Assumption 2.1, we have
\[ \frac{\hat{\mu}_E - \hat{\mu}_C - (\mu_E - \mu_C)}{\sqrt{\sigma_E^2/n_E + \sigma_C^2/n_C}} \to N(0,1) \]
in distribution as both \( n_E \to \infty \) and \( n_C \to \infty \). Therefore the power of the test under \( H_1 \) can be approximated by
\[ P\left(Z_1 > z_{(\alpha)} - \frac{\hat{\mu}_E - \hat{\mu}_C}{\sqrt{\sigma_E^2/n_E + \sigma_C^2/n_C}}\right) = 1 - \Psi\left(z_{(\alpha)} - \frac{\mu_E - \mu_C}{\sqrt{\sigma_E^2/n_E + \sigma_C^2/n_C}}\right), \]
where \( Z_1 \) is a standard normal random variable and \( \Psi \) is the cumulative function of the standard normal distribution.

From Assumption 2.2, we can replace \( n_E \) and \( n_C \) by \( vn + \tau \sqrt{n}Z \) and \( (1 - \nu)n - \tau \sqrt{n}Z \) respectively, where \( Z \) is a standard normal random variable. The approximate power is then
\[ \beta_0(n) = 1 - \Psi\left(z_{(\alpha)} - \frac{\mu_E - \mu_C}{\sqrt{\sigma_E^2/(vn + \tau \sqrt{n}Z) + \sigma_C^2/((1 - \nu)n - \tau \sqrt{n}Z)}}\right). \quad (6) \]
Thus, the mean power $\beta(n) = E(\beta_0(n))$ of a fixed $n$ is

$$1 - \int_{-\infty}^{\infty} \Psi \left( z(\alpha) - \frac{\mu_E - \mu_C}{\sqrt{\sigma_E^2/(\nu n + \tau \sqrt{n}) + \sigma_C^2/((1 - \nu) n - \tau \sqrt{n})}} \right) \psi(x) dx. \tag{7}$$

Where $\psi$ and $\Psi$ are the density and cumulative function of standard normal distribution.

To achieve a fixed power $\beta$ on average, we just find the smallest $n$ such that $\beta(n) \geq \beta$. From above derivation, we have the following Theorem.

**Theorem 2.1** Under Assumption 2.1 and 2.2, the sample size $n_1$ to achieve a fixed power $\beta$ on average is the smallest $n$ such that $\beta(n) \geq \beta$, where the power function $\beta(n)$ is defined in (7).

Sometimes, we want to find a sample size $n$, such that the power $\beta_0(n)$ is greater than or equal to $\beta$ with certain probability (0.90 is used in this paper). That is

$$\frac{\sigma_E^2}{\nu n + Z\tau \sqrt{n}} + \frac{\sigma_C^2}{(1 - \nu) n - Z\tau \sqrt{n}} < \left( \frac{\mu_E - \mu_C}{z(\alpha) + z(1-\beta)} \right)^2$$

at least with probability 0.90. Thus, we have the following theorem.

**Theorem 2.2** Under Assumptions 2.1 and 2.2, the sample size $n_2$, to achieve a fixed power $\beta$ with at least 90% confidence, is obtained to be the smallest integer, which satisfies

$$\frac{\sigma_E^2}{\nu n - 1.645\tau \sqrt{n}} + \frac{\sigma_C^2}{(1 - \nu) n + 1.645\tau \sqrt{n}} < \left( \frac{\mu_E - \mu_C}{z(\alpha) + z(1-\beta)} \right)^2 \tag{8}$$

and

$$\frac{\sigma_E^2}{\nu n + 1.645\tau \sqrt{n}} + \frac{\sigma_C^2}{(1 - \nu) n - 1.645\tau \sqrt{n}} < \left( \frac{\mu_E - \mu_C}{z(\alpha) + z(1-\beta)} \right)^2, \tag{9}$$

where $\tau$ is given in Assumption 2.2.

**Remark 2.4** From (6), we found that the power depends on the proportion $\nu$ and the variability $\tau$ of a randomization procedure. When $\nu$ is fixed, the power is a decreasing function of $\tau$. This has been demonstrated by simulation studies in Melfi and Page (1998) and Rosenberger, Stallard, Ivanova, Harper, and Ricks (2001). Further discussion of this power function is available in Section 5.

The results of Theorem 2.1 and 2.2 are based on asymptotic properties of randomization procedures. In the following two sections, we use these results to calculate the requisite sample sizes and study their finite sample properties. For simplicity of notation, we will use $n_0$ to represent the sample size from (5), $n_1$ to represent the sample size from Theorem 2.1 and $n_2$ to represent the sample size from Theorem 2.2.
3 Behrens-Fisher Problem

Assume that $X_1, X_2, \ldots$ and $Y_1, Y_2, \ldots$ are independent random variables with

$$X_1, X_2, \ldots \sim N(\mu_E, \sigma_E^2) \text{ and } Y_1, Y_2, \ldots \sim N(\mu_C, \sigma_C^2),$$

where the four parameters $\mu_E$, $\mu_C$, $\sigma_E^2$ and $\sigma_C^2$ are unknown. Let

$$\hat{\mu}_E = \bar{X} = n_E^{-1} \sum_{i=1}^{n_E} X_i, \quad \hat{\mu}_C = \bar{Y} = n_C^{-1} \sum_{i=1}^{n_C} Y_i, \quad \hat{\sigma}_E^2 = (n_E - 1)^{-1} \sum_{i=1}^{n_E} (X_i - \bar{X})^2$$

and

$$\hat{\sigma}_C^2 = (n_C - 1)^{-1} \sum_{i=1}^{n_C} (Y_i - \bar{Y})^2.$$

Consider three designs: (i) the complete randomization; (ii) the doubly adaptive biased coin design (DBCD) proposed by Eisele (1994), Eisele and Woodroofe (1995); and iii) the sequential maximum likelihood procedure (SMLE) by Melfi and Page (1998). These are each described in Rosenberger and Lachin (2002).

(i) **Complete randomization (CR):** Assign each patient to treatment group with $1/2$ probability. It is easy to see that

$$\frac{n_E}{n} \to \frac{1}{2} \text{ and } \sqrt{n} \left( \frac{n_E}{n} - .5 \right) \to N(0, \frac{1}{4})$$

in distribution. The sample size $n_E$ is estimated as the smallest integer, which is larger than

$$\frac{2(\sigma_E^2 + \sigma_C^2)(z(\alpha) + z(1-\beta))^2}{(\mu_E - \mu_C)^2}.$$  \hspace{1cm} (10)

The sample sizes $n_1$ and $n_2$ are defined in Theorem 2.1 and 2.2 respectively with $\nu = 1/2$ and $\tau = 1/2$.

(ii) **Doubly adaptive biased coin design (DBCD):** If the variances $\sigma_E^2$ and $\sigma_C^2$ are known, then the optimal allocation proportions for minimizing the total sample size and retaining preassigned coverage probability (Jennison and Turnbull (2000)) are

$$\frac{n_E}{n} \to \frac{\sigma_E}{\sigma_E + \sigma_C} \text{ and } \frac{n_C}{n} \to \frac{\sigma_C}{\sigma_E + \sigma_C}.$$
But $\sigma_E^2$ and $\sigma_C^2$ are usually unknown, Eisele (1994), Eisele and Woodroffe (1995) and Hu and Zhang (2002) propose the following doubly adaptive biased coin design. To start, allocate $m \geq 2$ patients to both treatment and control group. At any $(j + 1)$th stage ($j \geq 2m$), suppose there are $n_{E_j}$ observations on treatment and $n_{C_j}$ on control, with $n_{E_j} + n_{C_j} = j$. Let

$$X_j = n_{E_j}^{-1} \sum_{i=1}^{n_{E_j}} X_i, \quad \tilde{Y}_j = n_{C_j}^{-1} \sum_{i=1}^{n_{C_j}} Y_i, \quad \hat{\sigma}_{E_j}^2 = (n_{E_j} - 1)^{-1} \sum_{i=1}^{n_{E_j}} (X_i - \bar{X}_j)^2$$

and

$$\hat{\sigma}_{C_j}^2 = (n_{C_j} - 1)^{-1} \sum_{i=1}^{n_{C_j}} (Y_i - \bar{Y}_j)^2$$

be the estimates at stage $j + 1$. The estimated optimal allocation proportion is

$$\hat{\nu}_j = \frac{\hat{\sigma}_{E_j}}{\hat{\sigma}_{E_j} + \hat{\sigma}_{C_j}}.$$ 

The $(j + 1)$th patient is allocated to treatment with probability

$$q_{j+1} = q\left(\frac{n_{E_j}}{j}, \hat{\nu}_j\right),$$

where $q(\ldots)$ is a given allocation function. The following family of allocation is proposed by Hu and Zhang (2002):

$$q^{(\gamma)}(x, y) = \frac{y(x/y)^\gamma}{y(x/y)^\gamma + (1 - y)[(1 - y)/(1 - x)]^\gamma}, \quad q^{(\gamma)}(0, y) = 1, \quad q^{(\gamma)}(1, y) = 0,$$

with $\gamma \geq 0$.

From Hu and Zhang (2002), if $q^{(\gamma)}(x, y)$ is used in the doubly adaptive biased coin design, we have

$$\frac{n_{E}}{n} \rightarrow \nu = \frac{\sigma_{E}}{\sigma_{E} + \sigma_{C}}$$

and

$$\sqrt{n}\left(\frac{n_{E}}{n} - \frac{\sigma_{E}}{\sigma_{E} + \sigma_{C}}\right) \rightarrow N(0, \frac{5 + 3\gamma}{2(1 + 2\gamma)} \frac{\sigma_{E}\sigma_{C}}{(\sigma_{E} + \sigma_{C})^2})$$

in distribution. For fixed $\alpha$ and power $\beta$, $n_0$ is estimated as the smallest integer, which is larger than

$$\frac{(\sigma_{E} + \sigma_{C})^2(z_{(1-\alpha)} + z_{(1-\beta)})^2}{(\mu_{E} - \mu_{C})^2}.$$  \hspace{1cm} (11)
Table 1: Sample sizes of CR, SMLE and DBCD ($\alpha = .05$, $\beta = .8$ and $\mu_E - \mu_C = 1$).

<table>
<thead>
<tr>
<th></th>
<th>(1, 1)</th>
<th>(1, 2)</th>
<th>(1, 4)</th>
<th>(1, 8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n_0$ (CR)</td>
<td>25</td>
<td>62</td>
<td>211</td>
<td>804</td>
</tr>
<tr>
<td>$n_1$ (CR)</td>
<td>26</td>
<td>63</td>
<td>212</td>
<td>805</td>
</tr>
<tr>
<td>$n_2$ (CR)</td>
<td>28</td>
<td>72</td>
<td>234</td>
<td>853</td>
</tr>
<tr>
<td>$n_0$ (DBCD)</td>
<td>25</td>
<td>56</td>
<td>155</td>
<td>501</td>
</tr>
<tr>
<td>$n_1$ (SMLE)</td>
<td>28</td>
<td>59</td>
<td>158</td>
<td>504</td>
</tr>
<tr>
<td>$n_2$ (SMLE)</td>
<td>32</td>
<td>65</td>
<td>165</td>
<td>512</td>
</tr>
<tr>
<td>$n_1$ (DBCD, $\gamma = 1$)</td>
<td>27</td>
<td>58</td>
<td>157</td>
<td>503</td>
</tr>
<tr>
<td>$n_2$ (DBCD, $\gamma = 1$)</td>
<td>29</td>
<td>61</td>
<td>160</td>
<td>507</td>
</tr>
<tr>
<td>$n_1$ (DBCD, $\gamma = 4$)</td>
<td>26</td>
<td>57</td>
<td>156</td>
<td>502</td>
</tr>
<tr>
<td>$n_2$ (DBCD, $\gamma = 4$)</td>
<td>28</td>
<td>59</td>
<td>159</td>
<td>506</td>
</tr>
</tbody>
</table>

$n_1$ and $n_2$ are defined in Theorem 2.1 and 2.2 respectively with

$$\nu = \frac{\sigma_E}{\sigma_E + \sigma_C} \quad \text{and} \quad \tau = \frac{5 + 3\gamma}{2(1 + 2\gamma)} \frac{\sigma_E\sigma_C}{(\sigma_E + \sigma_C)^2}.$$  

(i) Sequential maximum likelihood procedure (SMLE): This method has been proposed by Melfi and Page (1998) in the context of Neyman allocation for binary responses. Hu and Zhang (2002) show that it is a special case of DBCD with Hu and Zhang’s allocation function ($\gamma = 0$). This is because when $\gamma = 0$, $q^0(x, y) = y$ and this leads to the SMLE.

Remark 3.1 Eisele (1994) use the allocation $q(x, y) = [1 - (1/y - 1)x]_+$. The disadvantages of this allocation function is discussed in Melfi, Page and Geraldes (2001). The allocation function $q^\gamma(x, y)$ proposed by Hu and Zhang (2002) depends on the parameter $\gamma$, which controls the randomness of the procedure, and it can be chosen to reflect the tradeoff between the degree of randomization and the variation. When $\gamma = 0$ (the SMLE), it has the largest variance $\tau$. When $\gamma = \infty$, the variance $\tau$ is minimized, but the procedure is complete predictable. Further asymptotic properties of the DBCD can be found in Hu and Zhang (2002).

For given $\alpha = 0.05$ and $\beta = 0.8$, Table 1 reports the required sample sizes ($n_0$, $n_1$ and $n_2$) for $\mu_E - \mu_C = 1$ and some different $\sigma_E$ and $\sigma_C$ values.
When $\sigma_E = \sigma_C$, the sample sizes $n_0$ of the three designs are the same. The DBCDs and SMLE have substantial advantages over the complete randomization when $\sigma_E$ is different from $\sigma_C$. For example, when $\sigma_E = 1$ and $\sigma_C = 4$, the sample size $n_0$ of DBCD is 501, which is significantly smaller than 804 of CR. So a well chosen response-adaptive randomization procedure can reduce sample size significantly in clinical trial.

The results in Table 1 show that $n_1$ is usually greater than $n_0$ by two or three. This is because when the value $n_E$ (random) shifts to one side, the power will decrease, but when it shifts to another, the power will increase. Therefore, it does not lose too much power on average comparing with a fixed design, when the variation of a randomization procedure is not very big. In this case, the sample size $n_1$ is not very different from $n_0$. But as we will find in next Section, $n_1$ could be very different from $n_0$, if the variation is very big.

The power of a randomization procedure is usually a random variable. In practice, we just run the experiment once, sometime we want to make sure the power will be achieved with certainly probability. Table 1 provides the sample sizes $n_2$ for this propose. We find that $n_2$ could much bigger than the ordinal sample size $n_0$. For example, the complete randomization with $\sigma_E = 1$ and $\sigma_C = 2$, $n_0 = 62$, but $n_2 = 72$ is required to achieve the fixed power (0.80) with probability 0.9. For this case, if 62 sample is used for a complete randomization, with 20\% of the chance, the power is less than 0.76, when the target power is 0.80. This is due to the variation of CR. This disadvantage was pointed in Efron (1971). Similarly, for the SMLE, $n_0 = 56$, but $n_2 = 65$. The DBCDs are proposed to balance between randomness and variation. From Table 1, we can find that $n_2$ is slight larger than $n_0$ for DBCD with $\gamma = 4$. Therefore, without substantial change of sample size, a well planned randomization procedure can still achieve a fixed power with high probability.

Sample sizes in Table 1 are based on the power function (7), which depends on the asymptotic distribution of $n_E$ and $n_C$. It is important to know their finite properties. To see this, we just have to compare the approximate power function (7) with its simulated power function. We consider the case with $\mu_E = 1$, $\mu_C = 0$, $\sigma_E = 1$, $\sigma_C = 2$, $\alpha = .95$ and $\beta = .80$. Each simulation is done as the doubly adaptive biased coin design with $m = 5$. From $n = 50$ to $n = 70$, we use the test (1) to record reject or accept. In Figure 1, the solid line represents the power function based on (7) for SMLE (or DBCD with $\gamma = 0$). The dots are the average powers of the 10000 simulations for different sample sizes from 50 to 70 for SMLE. In Figure 2, the simulated and approximate powers of DBCD ($\gamma = 4$) is reported.

From these two figures, the power function (7) provides a good approximation of the real power. So the sample size based on this power function is also reliable.
4 Response-adaptive randomization procedures for binary response trials

Response-adaptive randomization procedures for clinical trials with binary responses have been widely studied in the literature (Flournoy and Rosenberger (1995), Rosenberger and Lachin (2002), Chapters 10-12). Suppose $X_1, X_2, \ldots, X_{nE}$ represent the responses in the experimental group and $Y_1, Y_2, \ldots, Y_{nC}$ are the observations from the controlled treatment. Let $p_E = P(X_1 = 1)$, $p_C = P(Y_1 = 1)$, $q_E = 1 - p_E$ and $q_C = 1 - p_C$. The estimators are $\hat{p}_E = \bar{X} = \sum_{i=1}^{n_E} X_i / n_E$, $\hat{p}_C = \bar{Y} = \sum_{i=1}^{n_C} Y_i / n_C$, $\hat{\sigma}_E^2 = \hat{p}_E(1 - \hat{p}_E)/n_E$ and $\hat{\sigma}_C^2 = \hat{p}_C(1 - \hat{p}_C)/n_C$. Three randomization procedures are considered in this section.

(i) Randomized play-the-winner (RPW) rule: One of the most widely studied adaptive design is the randomized play-the-winner (RPW) rule (Wei and Durham, 1978). We start with $2a$ balls in the urn with $a$ type E balls and $a$ type C balls. When a patient is ready to be randomized, a ball is drawn and replaced. If the ball is type E (C), the patient is assigned to experimental (controlled) treatment. A success on the treatment generates a ball of the same type to the urn. A failure on the treatment generates a ball of the opposite type to the urn.

Athreya and Karlin (1968) showed that

$$\frac{n_E}{n_E + n_C} \to \frac{q_C}{q_E + q_C} \text{ almost surely as } n \to \infty. \quad (12)$$

When $p_E + p_C < 1.5$ (or $q_E + q_C > 0.5$), we have the following asymptotic normality of $N_E$:

$$\sqrt{n}(\frac{n_E}{n_E + n_C} - \frac{q_C}{q_E + q_C}) \to N(0, \sigma^2), \quad (13)$$

in distribution, where

$$\sigma^2 = \frac{[5 - 2(q_E + q_C)]q_EQC}{[2(q_E + q_C) - 1][q_E + q_C]^2},$$

(cf. Smythe and Rosenberger, 1995). When $p_E + p_C > 1.5$, the limiting distributions of the proportion of patients assigned to treatment are unknown.

The sample size $n_0$ can be calculated as the smallest integer, which is larger than

$$\frac{(q_E + q_C)(p_0 + p_Cq_C)(z_\alpha + z_{1-\beta})^2}{(p_E - p_C)^2}.$$
$n_1$ and $n_2$ are calculated from Theorem 2.1 and 2.2 respectively with

$$\nu = \frac{q_c}{q_E + q_C} \text{ and } \tau^2 = \frac{[5 - 2(q_E + q_C)]q_Eq_C}{[2(q_E + q_C) - 1](q_E + q_C)^2}.$$  

(ii) **Doubly adaptive biased coin design:** Suppose we target the same allocation proportion as the RPW rule. The design is following. To start, allocate $m$ patients to both treatment and control group. At any $(j+1)$th stage $(j \geq 2m)$, suppose there are $n_{Ej}$ observations on treatment and $n_{Cj}$ on control, with $n_{Ej} + n_{Cj} = j$. Let

$$\hat{p}_{Ej} = n_{Ej}^{-1} \sum_{i=1}^{n_{Ej}} X_i, \text{ and } \hat{p}_{Cj} = n_{Cj}^{-1} \sum_{i=1}^{n_{Cj}} Y_i$$

The estimated allocation proportion is now

$$\hat{p}_j = \frac{1 - \hat{p}_{Cj}}{2 - \hat{p}_{Ej} - \hat{p}_{Cj}}.$$ 

The $(j + 1)$th patient is allocated to treatment with probability

$$q_{j+1} = q\left(\frac{n_{Ej}}{j}, \hat{p}_j\right),$$

where $q(., .)$ is a given allocation function. If the $q_{(\gamma)}(x, y)$ of Section 3 is used, it can be shown (Hu and Zhang, 2002) that

$$\frac{n_E}{n} \rightarrow \nu = \frac{q_c}{q_E + q_C}$$

and

$$\sqrt{n}\left(\frac{n_E}{n} - \frac{q_c}{q_E + q_C}\right) \rightarrow N\left(0, \frac{q_Eq_C(p_E + p_C)}{(q_E + q_C)^3} + \frac{2q_Eq_C}{(1 + 2\gamma)(q_E + q_C)^3}\right)$$

in distribution. For fixed $\alpha$ and power $\beta$, the sample size $n_0$ is same as RPW rule. The sample size $n_1$ and $n_2$ are defined by Theorem 2.1 and 2.2 respectively with

$$\nu = \frac{q_c}{q_E + q_C} \text{ and } \tau^2 = \frac{q_Eq_C(p_E + p_C)}{(q_E + q_C)^3} + \frac{2q_Eq_C}{(1 + 2\gamma)(q_E + q_C)^3}).$$

When $p_E = .8$ and $p_C = .7$, the $n_1$ and $n_2$ are not available for RPW rule, this is because the variance of RPW rule does not exist for $p_E + p_C \geq 1.5$. When $p_E + p_C$ is not much less than 1.5, the variance of RPW rule could be very large, this is reflected
in the sample sizes \( n_1 \) and \( n_2 \) for the case of \( p_E = .8 \) and \( p_C = .6 \). The DBCDs have much smaller variances than the RPW rule in those cases. When both \( p_E \) and \( p_C \) are small, the RPW rule and the DBCDs have similar variances, so \( n_1 \) and \( n_2 \) are similar.

When both \( q_E \) and \( q_C \) are very small, the variance \( \tau^2 \) of DBCD could be very large, no matter what the value \( \gamma \) is used. This is because the first term of \( \tau^2 \) does not depend on \( \gamma \) and converges to \( \infty \) at both \( q_E \) and \( q_C \) converge to 0. In these cases, \( n_1 \) and \( n_2 \) are very different from \( n_0 \).

To see finite sample properties of the power function (7), we consider the case with \( p_E = 0.3 \) and \( p_C = 0.2 \) for both RPW and DBCD (\( \gamma = 4 \)). For the RPW procedures, we start with 2 balls of each type. For the DBCDs, we start with \( m = 5 \). Figure 3 is based on the 10000 simulations of RPW rule. Again the solid line is the approximate power function from (7) and the dots are the simulated powers. The simulated and approximate powers of DBCD (\( \gamma = 4 \)) is reported in Figure 4. These figures show that the (7) provides a good approximation of the power.

## 5 Discussion

From the equations of \( \beta_0(n) \) and \( \beta(n) \), the power depends on both allocation proportion \( \nu \) and variance \( \tau^2 \). Hayre (1979), Jennison and Turnbull (2000), Rosenberger, Stallard, Ivanova, Harper, and Ricks (2001), among others, have discussed some optimization criteria to choose \( \nu \). For the binary response trials in Section 4, 
\[
\nu = \sqrt{p_E q_E / (p_E q_E + p_C q_C)} \quad \text{Neyman allocation}
\]
maximizes the power \( \beta_0(n) \). Melfi, Page and Geraldes (2001) use SMLE procedure to target the Neyman allocation.
\( \nu = \sqrt{p_E/(p_E + p_C)} \) is proposed in Rosenberger, et al (2001). In their paper, SMLE procedure is used to target \( \nu = \sqrt{p_E/(p_E + p_C)} \). But in these papers, sample sizes are not calculated. By using the results in Section 2, we can estimate the required sample sizes for a given power.

When \( \nu \) is fixed, both \( \beta_0(n) \) and \( \beta(n) \) are decreasing function of \( \tau^2 \). From Section 3 and 4, we see that the DBCD (with \( \gamma > 0 \)) always has smaller variance than the SMLE procedure. So we can use the DBCD to target the above \( \nu \)'s and to achieve larger power (or reduce the sample size). The RPW rule cannot be used to target the above \( \nu \)'s. Also the RPW rule only can be used for binary response trials. From the studies of Section 3 and 4, the DBCD is the most favorable procedure in this paper. Therefore, \( \beta_0(n) \) and \( \beta(n) \) can also be used to choose allocation proportion \( \nu \) and to evaluate a randomization procedure.

In this paper, we have only consider sample sizes of randomization procedures with two treatments. Response-adaptive randomization procedures for \( K \) treatments have been proposed and studied extensively recently. Urn models for \( K \) treatments have been explored by Wei (1979), Smythe (1996), Durham, Flournoy, and Li (1998), and Bai, Hu and Shen (2001). Some general asymptotic properties have been developed by Bai and Hu (1999). \( K \) treatments adaptive biased coin designs and their properties have been studied in Smith (1984) and Wei, Smythe and Smith (1986). A DBCD of \( K \) treatments has been proposed and studied in Hu and Zhang (2002). In that paper, asymptotic normality is obtained. But it remains unknown about how to calculate sample sizes of randomization procedures with \( K > 2 \) treatments.

Typically, clinical trials do not result in immediate outcomes. For urn models, we can update the urn when the outcomes become available. This is first suggested by Wei (1988) for the RPW rule. Bai, Hu and Rosenberger (2002) study asymptotic properties of urn models with delayed response. From that paper, the asymptotic properties of the urn structure do not change with delayed response. The asymptotic distribution of allocations is not available in that paper. For the DBCD, delayed response will effect the estimators of the unknown parameters. The asymptotic properties of the DBCD with delayed response are unknown. For both procedures, it is still unclear how the sample sizes will be effect by delayed response.

The main contribution of this paper is in deriving sample sizes of a randomization procedure by taking account of its variation. We have also demonstrated the relationship among the power, the allocation proportion, the variation and the sample size of a randomization procedure with two treatments.
References


