

# Optimal biased coins for two-arm clinical trials

THOMAS E. GWISE, JIANHUA HU AND FEIFANG HU

---

Atkinson (1982) introduced  $D$ -optimal and  $D_A$ -optimal biased coin designs to achieve balanced allocation. In this paper, we relax the restrictive assumptions on Atkinson's design so that it can be applied to more realistic situations of heteroscedastic data for two-arm clinic trials. A new class of adaptive biased coin designs is obtained. Further, we derive the asymptotic properties of the proposed allocation procedures, and show that the allocation targets maximizing power via a direct connection to Neyman's allocation scheme. Simulation studies illustrate that the theoretical results are valid and that the proposed design has testing power that is superior to that of the completely randomized allocation scheme.

KEYWORDS AND PHRASES: Adaptive design,  $D$ -optimal biased coin,  $D_A$ -optimal biased coin, Response adaptive design, Continuous responses, Binary responses.

---

## 1. INTRODUCTION

In two-arm clinical trials, patients arrive sequentially and are assigned to one of two treatments. The objective is to assess the performance of the two treatments. Traditionally, a randomized design is employed in two-arm clinical trials, which assigns each patient to one of the two treatments with a 50:50 probability. While providing the advantage of reducing selection bias, this allocation scheme, however, carries potential problems such as an imbalanced allocation and therefore may unnecessarily allocate more patients to the inferior treatment. To achieve allocation balance, Efron (1971) introduced biased coin designs using a measure of imbalance to determine the treatment to which the next patient would be assigned with higher probability. Wei (1977, 1978) developed an urn model to assign patients probabilistically by changing the probability of the next patient assignment to the under-represented treatment gradually according to the amount of imbalance.

Another large family of allocation schemes, the response adaptive design, has been increasingly used in clinical trials, and can be used to target several objectives, i.e., a frequently targeted objective is to maximize the power of the test. Among several proposed methods, the urn model (Athreya and Karlin 1968, Wei and Durham 1978, Rosenberger 2002) represents a broad category that uses response information

from the ongoing clinical trial to adjust the number of different colored balls (representing treatments) in the urn. Another class of response-adaptive randomized allocation is the optimal target allocation design, which must first choose a target (Rosenberger and Lachin 2002, Jennison and Turnbull 2000). The class of the doubly adaptive biased coin design (DBCD) has also been discussed in the literature (Eisele 1994, Eisele and Woodroffe 1995, Hu and Zhang 2004). It targets an allocation ratio while monitoring the current allocation ratio. A general classification of adaptive designs is defined and discussed in Hu and Rosenberger (2006).

Atkinson (1982) proposed what we are calling the  $D$ -optimal and  $D_A$ -optimal biased coin designs for the purpose of balancing the patients with respect to treatment assignments. Atkinson argues that this type of design not only combats selection bias, but also eliminates the need to use arbitrary functions to achieve the desired balance in patient allocation (i.e., Efron's biased coin). A key assumption of Atkinson's design is the linear modeling of treatment responses with a common variance for different treatments. The  $D$ -optimal design assigns treatments to subjects such that the determinant of the covariance matrix is minimized. The  $D_A$ -optimal design is employed to target some interesting and estimable functions of the treatment effects (i.e., the difference in treatment effects). An important drawback of Atkinson's design is the strong assumption of homoscedastic distribution of the responses with respect to treatments. In many real applications, the variances are heteroscedastic among different treatments. See Example 2 or Hu and Rosenberger (2006) for more examples. Also asymptotic properties of Atkinson's designs are not available in the literature. As pointed out in Hu and Rosenberger (2003), the asymptotic distribution of the allocation proportion is very important in clinical investigations.

The main objectives are (i) to propose a new class of adaptive design,  $D$ -optimal and  $D_A$ -optimal biased coin design, for heteroscedastic outcomes with both known and unknown variances, and (ii) to study asymptotic and numerical properties of the proposed designs. These properties provide a solid foundation for these optimal biased coin designs and their statistical inferences. Asymptotic properties of Atkinson's designs can then be obtained as a special case.

The paper is organized as follows. In Section 2, we propose the generalized  $D$ -optimal and  $D_A$ -optimal biased coin designs. Some asymptotic results (including the strong consistency and asymptotic normality) of both the estimators of unknown parameters and the allocation proportions are

derived under widely satisfied conditions. The proofs are sketched in the Appendix. In Section 3, we outline the theoretical results in two practical examples using both designs. Section 4 describes some simulation studies to verify the derived theoretical results and make a comparison between the proposed methods and 50:50 randomized allocation in terms of power. We present a conclusion and some discussion in Section 5.

## 2. OPTIMAL DESIGNS AND MAIN RESULTS

### 2.1 Generalized $D$ -optimal biased coin designs

The  $D$ -optimal biased coin design in Atkinson (1982) was formulated based on the  $D$ -optimal design in the context of linear models

$$(2.1) \quad Y_j = \mathbf{X}'_j \boldsymbol{\beta} + e_j,$$

where  $e_j$  ( $j = 1, \dots, n$ ) are independent random errors with mean 0 and variance  $\sigma^2$ . The variance of the least squares estimate of  $\boldsymbol{\beta}$  is then

$$\text{Cov}(\hat{\boldsymbol{\beta}}) = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1},$$

where  $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_n)$ . In the problem of allocations of two treatments, we define  $\mathbf{M}(\xi) = \mathbf{X}'\mathbf{X}$ , where  $\xi$  is a measure over the design region  $\mathcal{X} = \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}$ . Here  $\mathbf{x} = (1, 0)'$  represents treatment 1 and  $\mathbf{x} = (0, 1)'$  represents treatment 2. The  $D$ -optimal design aims to minimize the determinant of the generalized variance  $\mathbf{M}^{-1}(\xi)$ . Another slightly different  $G$ -optimal design chooses  $\mathbf{x}$  to minimize  $\sup_{\mathbf{x} \in \mathcal{X}} d(\mathbf{x}, \xi)$ , where the standardized variance  $d(\mathbf{x}, \xi) = \mathbf{x}' \mathbf{M}^{-1}(\xi) \mathbf{x}$ . Keifer and Wolfowitz (1960) showed in the general equivalence theorem that  $D$ -optimality and  $G$ -optimality designs are equivalent for continuous design measures  $\xi$ . Atkinson shows that the equivalence of  $D$ -optimal and  $G$ -optimal designs provides a useful tool for sequentially constructing  $D$ -optimal designs. He proposed combining  $D$ -optimality with a biased coin to reduce selection bias that could be caused by the deterministic nature of assigning patients sequentially to the treatment that maximizes  $d(\mathbf{x}_i, \xi) = \mathbf{x}'_i (\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}_i$ ,  $i = 1, 2$ , where  $\mathbf{x}_1 = (1, 0)'$  and  $\mathbf{x}_2 = (0, 1)'$ .

According to Atkinson's design, we let  $P_{l,i}$  be the probability of the  $l$ -th patient assigned to treatment  $i$  and  $n_{l,i}$  be the number of patients assigned to treatment  $i$  after  $l$  patients have been enrolled. The  $(l+1)$ st patient is assigned to treatment 1 with the probability

$$(2.2) \quad P_{l+1,1} = \frac{d(\mathbf{x}_1, \xi)}{d(\mathbf{x}_1, \xi) + d(\mathbf{x}_2, \xi)} = \frac{n_{l,2}}{n_{l,1} + n_{l,2}}.$$

Note that the procedure requires some initial number of patients to be randomized prior to implementing the biased coin. As observed in Atkinson (1982), allocation in (2.2) forces balance among two treatments as Efron (1971).

Atkinson's model in (2.1) assumes the outcomes are homogeneous among the two treatments. Here we relax this assumption to

$$(2.3) \quad \text{Var}(e_j | \mathbf{X}_j = (1, 0)') = \sigma_1^2 \quad \text{and} \quad \text{Var}(e_j | \mathbf{X}_j = (0, 1)') = \sigma_2^2.$$

In this case, the least squared estimate of  $\boldsymbol{\beta}$  has

$$(2.4) \quad \text{cov}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \Sigma_y \mathbf{X} (\mathbf{X}'\mathbf{X})^{-1} \\ = \begin{pmatrix} \sigma_1^2/n_{l,1} & 0 \\ 0 & \sigma_2^2/n_{l,2} \end{pmatrix}.$$

Following a procedure similar to Atkinson's approach, the probability of assigning the  $(l+1)$ st patient to treatment 1 is

$$(2.5) \quad P_{l+1,1} = \frac{d(\mathbf{x}_1, \xi)}{d(\mathbf{x}_1, \xi) + d(\mathbf{x}_2, \xi)} = \frac{\sigma_1^2/n_{l,1}}{\sigma_1^2/n_{l,1} + \sigma_2^2/n_{l,2}}.$$

In practical cases where the variances are unknown, we substitute estimators for the unknown variances. Let  $\hat{\sigma}_{l,i}^2$  be an estimate of  $\sigma_i^2$  under the  $i$ -th treatment after  $l$  responses have been observed. The probability of assigning the  $(l+1)$ -st patients to treatment 1 is

$$(2.6) \quad P_{l+1,1} = \frac{\hat{\sigma}_{l,1}^2/n_{l,1}}{\hat{\sigma}_{l,1}^2/n_{l,1} + \hat{\sigma}_{l,2}^2/n_{l,2}}.$$

Note that the response information from the study is used to calculate subsequent assignment probabilities.

The design in (2.6) is based on  $D$ -optimality, which is different from the two main families in the literature (see Hu and Rosenberger (2006)): urn models and the procedures based on sequential estimation. As pointed out in Hu and Rosenberger (2003), it is important to have the asymptotic properties (especially the asymptotic limit and asymptotic variance) of the proportion of allocation. To state the asymptotic properties, we need following notations and assumptions.

Suppose that  $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)'$  are  $2d$ -dimension parameters in the model (2.3). Here  $\boldsymbol{\theta}_i$  ( $i = 1, 2$ ) is the unknown parameter corresponding to treatment  $i$ .  $\hat{\boldsymbol{\theta}}_l = (\hat{\boldsymbol{\theta}}_{l,1}, \hat{\boldsymbol{\theta}}_{l,2})'$  is an estimator of  $\boldsymbol{\theta}$  based on the first  $l$  responses. Let  $\mathbf{X}_j = (X_{j,1}, X_{j,2})'$  ( $j = 1, \dots, l$ ) be the treatment assignment. Let

$$n_{l,1} = \sum_{j=1}^l X_{j,1} \quad \text{and} \quad n_{l,2} = \sum_{j=1}^l X_{j,2}.$$

CONDITION 1. Suppose that for  $i = 1, 2$ ,

$$(2.7) \quad \hat{\boldsymbol{\theta}}_{l,i} - \boldsymbol{\theta}_i = \frac{1}{n_{l,i}} \sum_{j=1}^l X_{j,i} \mathbf{h}_i(Y_j, \boldsymbol{\theta}) + o(n_{l,i}^{-1/2}) \quad \text{almost surely,}$$

as  $n_{l,i} \rightarrow \infty$ . Where  $\mathbf{h}_i$  are  $d$  functions satisfy  $E\mathbf{h}_i(Y_j, \boldsymbol{\theta}) = 0$  and  $E\|\mathbf{h}_i(Y_j, \boldsymbol{\theta})\|^2 < \infty$ . Let  $\mathbf{V}_i = E\mathbf{h}_i(Y_j, \boldsymbol{\theta})(\mathbf{h}_i(Y_j, \boldsymbol{\theta}))'$  be the covariance matrices.

**CONDITION 2.** Suppose both  $\sigma_1(\boldsymbol{\theta})$  and  $\sigma_2(\boldsymbol{\theta})$  are twice continuously differentiable functions of  $\boldsymbol{\theta}$  in a small neighborhood of the true parameter. Also  $\sigma_1(\boldsymbol{\theta}) \neq 0$  and  $\sigma_2(\boldsymbol{\theta}) \neq 0$

**Remark 2.1.** In some applications,  $\boldsymbol{\theta} = (\beta_1, \sigma_1^2, \beta_2, \sigma_2^2)'$  as in Example 1 of Section 3. However, in some applications,  $\sigma_1^2$  and  $\sigma_2^2$  are related with  $\beta_1$  and  $\beta_2$  (as in Example 2), then  $\boldsymbol{\theta} = (\beta_1, \beta_2)'$  or some other forms. Here we introduce the new parameter  $\boldsymbol{\theta}$  to include all these cases.

**Remark 2.2.** Condition 1 is about the estimator  $\hat{\boldsymbol{\theta}}$ . When the estimator is a maximum likelihood estimator, moment estimator, or an estimator from estimating equations, this condition is usually satisfied. We will illustrate this in Section 3. Condition 2 is also widely satisfied in applications.

The asymptotic properties of the allocation procedure are stated in Theorem 1.

**Theorem 1.** Under Conditions 1 and 2, the  $D$ -optimal biased coin allocation procedure has the following properties as  $l \rightarrow \infty$ . (i) Strong consistency:

$$\boldsymbol{\theta}_l \rightarrow \boldsymbol{\theta} \quad \text{and} \quad \frac{n_{l,1}}{l} \rightarrow \frac{\sigma_1}{\sigma_1 + \sigma_2} \quad \text{almost surely.}$$

(ii) The law of iterated logarithm:

$$\boldsymbol{\theta}_l - \boldsymbol{\theta} = \left( \sqrt{\frac{\log \log(l)}{l}} \right)$$

and

$$\frac{n_{l,1}}{l} - \frac{\sigma_1}{\sigma_1 + \sigma_2} = O\left(\sqrt{\frac{\log \log(l)}{l}}\right) \quad \text{almost surely.}$$

And (iii) the asymptotic normality:

$$(2.8) \quad \sqrt{l}(\hat{\boldsymbol{\theta}}_l - \boldsymbol{\theta}) \rightarrow N(\mathbf{0}, \mathbf{V})$$

in distribution, where

$$\mathbf{V} = \text{diag}\left(\frac{\sigma_1}{\sigma_1 + \sigma_2} \mathbf{V}_1, \frac{\sigma_2}{\sigma_1 + \sigma_2} \mathbf{V}_2\right),$$

a positive definite matrix. Further, we have

$$(2.9) \quad \sqrt{l}\left(\frac{n_{l,1}}{l} - \frac{\sigma_1}{\sigma_1 + \sigma_2}\right) \rightarrow N\left(\mathbf{0}, \frac{1}{3} \frac{\sigma_1 \sigma_2}{(\sigma_1 + \sigma_2)^2} + \frac{4}{3} \tau^2\right)$$

in distribution, where

$$\tau^2 = (\nabla(\rho)|_{\boldsymbol{\theta}})' \mathbf{V} (\nabla(\rho)|_{\boldsymbol{\theta}}).$$

The proofs are outlined in the Appendix. If the outcome variances are known constants, then  $\rho(\boldsymbol{\theta})$  is constant with partial derivatives 0 and  $\nabla \rho(\boldsymbol{\theta}) = 0 \Rightarrow \tau^2 = 0$ . The extended results follows easily in Corollary 1.

**Corollary 1.** If the conditions for Theorem 1 are met and  $\sigma_i$  ( $i = 1, 2$ ) are known, then

$$(2.10) \quad \sqrt{l}\left(\frac{n_{l,1}}{l} - \nu_1\right) \xrightarrow{D} \mathcal{N}\left(\mathbf{0}, \frac{1}{3} \frac{\sigma_1 \sigma_2}{(\sigma_1 + \sigma_2)^2}\right).$$

When  $\sigma_1 = \sigma_2$  (Atkinson's  $D$ -optimal design), then

$$(2.11) \quad \sqrt{l}\left(\frac{n_{l,1}}{l} - \frac{1}{2}\right) \xrightarrow{D} \mathcal{N}\left(\mathbf{0}, \frac{1}{12}\right).$$

**Remark 2.3.** Based on Corollary 1, the asymptotic variance of Atkinson's  $D$ -optimal design is  $1/12$ . On the other hand, it can be shown that the asymptotic variance of the allocation probability using Wei's adaptive biased coin design targeting the balanced allocation (Wei 1977, 1978; Rosenberger and Lachin 2002) is no smaller than  $1/12$ , which only occurs in an unusual case. The result implies that Atkinson's  $D$ -optimal design performs better than Wei's urn design in terms of testing power (see Hu and Rosenberger, 2003). This agrees with the numerical results in Atkinson (1982).

## 2.2 $D_A$ -optimal biased coin design

The  $D_A$ -optimal biased coin design has been proposed to target the contrast between treatment effects, which are often of primary interest. Let  $A$  represent some estimable function. In two-arm treatment comparisons, typically  $A' = (1, -1)$ . The covariance matrix of  $A'\beta$  is proportional to  $A'M^{-1}A$ . The  $D_A$ -optimality design aims to maximize  $\det[A'M^{-1}A]^{-1}$ , analogous to  $D$ -optimality. Proceeding in a similar manner to that used in developing the  $D$ -optimal design, we define the following expression

$$(2.12) \quad d_A(\mathbf{x}, \xi) = \mathbf{x}' M^{-1}(\xi) A \{A' M^{-1}(\xi) A\}^{-1} A' M^{-1}(\xi) \mathbf{x}.$$

The above can be simplified to

$$d_A(\mathbf{x}, \xi) = \left(\frac{\sigma_1^2}{n_{l,1}} + \frac{\sigma_2^2}{n_{l,2}}\right)^{-1} \times \mathbf{x}' \begin{pmatrix} (\sigma_1^2/n_{l,1})^2 & -(\sigma_1^2 \sigma_2^2)/(n_{l,1} n_{l,2}) \\ -(\sigma_1^2 \sigma_2^2)/(n_{l,1} n_{l,2}) & (\sigma_2^2/n_{l,2})^2 \end{pmatrix} \mathbf{x}.$$

Then

$$(2.13) \quad P_{l+1,1} = \frac{d(\mathbf{x}_1, \xi)}{d(\mathbf{x}_1, \xi) + d(\mathbf{x}_2, \xi)} = \frac{(\sigma_1^2/n_{l,1})^2}{(\sigma_1^2/n_{l,1})^2 + (\sigma_2^2/n_{l,2})^2}.$$

Given the scenario where outcome variances are unknown, we use their estimators to replace them. The allocation prob-

ability follows as

$$(2.14) \quad P_{l+1,1} = \frac{(\hat{\sigma}_{l,1}^2/n_{l,1})^2}{(\hat{\sigma}_{l,1}^2/n_{l,1})^2 + (\hat{\sigma}_{l,2}^2/n_{l,2})^2}.$$

The asymptotic properties of the allocation scheme are stated in Theorem 2.

**Theorem 2.** *Under Conditions 1 and 2, the  $D_A$ -optimal biased coin allocation procedure has the following properties as  $l \rightarrow \infty$ . (i) Strong consistency:*

$$\theta_l \rightarrow \theta \quad \text{and} \quad \frac{n_{l,1}}{l} \rightarrow \frac{\sigma_1^{4/3}}{\sigma_1^{4/3} + \sigma_2^{4/3}} \text{ almost surely.}$$

(ii) *The law of iterated logarithm:*

$$\theta_l - \theta = \left( \sqrt{\frac{\log \log(l)}{l}} \right)$$

and

$$\frac{n_{l,1}}{l} - \frac{\sigma_1^{4/3}}{\sigma_1^{4/3} + \sigma_2^{4/3}} = O\left(\sqrt{\frac{\log \log(l)}{l}}\right) \text{ almost surely.}$$

And (iii) *the asymptotical normality:*

$$(2.15) \quad \sqrt{l}(\hat{\theta}_l - \theta) \rightarrow N(\mathbf{0}, \mathbf{W})$$

in distribution, where

$$\mathbf{W} = \text{diag}\left(\frac{\sigma_1^{4/3}}{\sigma_1^{4/3} + \sigma_2^{4/3}} \mathbf{V}_1, \frac{\sigma_2^{4/3}}{\sigma_1^{4/3} + \sigma_2^{4/3}} \mathbf{V}_2\right),$$

a positive definite matrix. Further, we have

$$(2.16) \quad \sqrt{l} \left( \frac{n_{l,1}}{l} - \frac{\sigma_1^{4/3}}{\sigma_1^{4/3} + \sigma_2^{4/3}} \right) \rightarrow N\left(\mathbf{0}, \frac{1}{5} \frac{(\sigma_1 \sigma_2)^{4/3}}{(\sigma_1^{4/3} + \sigma_2^{4/3})^2} + \frac{6}{5} \tau_1^2\right)$$

in distribution, where

$$\tau_1^2 = (\nabla(\rho)|_{\theta})' \mathbf{W} (\nabla(\rho)|_{\theta}).$$

The sketch of the proof is shown in the Appendix. If  $\sigma_1 = \sigma_2$  or the outcome variances are known constants, we have

**Corollary 2.** *If the conditions for Theorem 2 are met and  $\sigma_i$ ,  $i = 1, 2$  are known, then*

$$(2.17) \quad \sqrt{l} \left( \frac{n_{l,1}}{l} - \nu_1 \right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\mathbf{0}, \frac{1}{5} \frac{\sigma_1^{4/3} \sigma_2^{4/3}}{(\sigma_1^{4/3} + \sigma_2^{4/3})^2}\right).$$

When  $\sigma_1 = \sigma_2$  (Atkinson's  $D_A$ -optimal design), then

$$(2.18) \quad \sqrt{l} \left( \frac{n_{l,1}}{l} - \frac{1}{2} \right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\mathbf{0}, \frac{1}{20}\right).$$

**Remark 2.4.** Based on Corollary 1, the asymptotic variance of Atkinson's  $D_A$ -optimal design is  $1/20$ . The result implies that Atkinson's  $D_A$ -optimal design performs better than both Wei's urn designs and the  $D$ -optimal design in terms of testing power. This agrees with the objective of the  $D_A$ -optimal design that targets the contrasts between treatment effects.

**Remark 2.5.** From Theorem 1 and 2, both the limiting allocation proportions of  $D$ -optimal biased coin design and  $D_A$ -optimal biased coin design depend on the variances of the two treatments. Both designs assign more subjects to the treatment with higher variance. It is possible that both designs can assign more subjects to the inferior treatment group. This is because both  $D$ -optimal and  $D_A$ -optimal designs only emphasizing the treatment difference. Recently, Tymofyeyev, Rosenberger and Hu (2007) proposed some designs (not  $D$ -optimal or  $D_A$ -optimal) that minimize the number of patients into the inferior treatment group for fixed power. How to combine the idea of  $D$ -optimal or  $D_A$ -optimal with minimizing the number of patients into the inferior treatment group remains a further research topic.

### 3. SOME EXAMPLES

In this section, we use two examples to illustrate how to apply the Theorems in Section 2. We derive the asymptotic properties of continuous and dichotomous outcomes, the two typical cases in practice, using the  $D$ -optimal and  $D_A$ -optimal allocation schemes.

**EXAMPLE 1 (CONTINUOUS RESPONSES).** In a clinical trial, outcomes of patients on treatments  $i$  ( $i = 1, 2$ ),  $Y_{j,i}$ , are independently distributed with mean  $\mu_i$  and variance  $\sigma_i^2$ . This can be viewed as special case of model (2.3). The parameter  $\theta = (\mu_1, \sigma_1^2, \mu_2, \sigma_2^2)'$  in this case. Based on the  $l$  observations ( $n_{l,1}$  in treatment 1 and  $n_{l,2} = l - n_{l,1}$  in treatment 2), the variance estimates are

$$\hat{\sigma}_{l,1}^2 = (n_{l,1} - 1)^{-1} \sum_{i=1}^{n_{l,1}} (Y_{i,1} - \bar{Y}_{l,1})^2 \quad \text{and}$$

$$\hat{\sigma}_{l,2}^2 = (n_{l,2} - 1)^{-1} \sum_{i=1}^{n_{l,2}} (Y_{i,2} - \bar{Y}_{l,2})^2,$$

where  $\bar{Y}_{l,1} = n_{l,1}^{-1} \sum_{i=1}^{n_{l,1}} Y_{i,1}$  and  $\bar{Y}_{l,2} = n_{l,2}^{-1} \sum_{i=1}^{n_{l,2}} Y_{i,2}$ . The  $D$ -optimal biased coin design is then assigning the  $(l+1)$ st patient to treatment 1 with probability as (2.6). The  $D_A$ -optimal biased coin design is assigning the  $(l+1)$ st patient to treatment 1 with probability (2.14).

Based on Theorem 1 and Theorem 2, we have the following result.

**Corollary 3.** *Suppose that  $\gamma_1 = E(e_j^4 | \mathbf{X}_j = (1, 0)') < \infty$  and  $\gamma_2 = E(e_j^4 | \mathbf{X}_j = (0, 1)') < \infty$ . Then*

$$(3.1) \quad \sqrt{l} \left( \frac{n_{l,1}}{l} - \frac{\sigma_1}{\sigma_1 + \sigma_2} \right) \rightarrow N(0, \eta^2),$$

where

$$\eta^2 = \frac{\sigma_1\sigma_2}{3(\sigma_1 + \sigma_2)^2} + \frac{(\gamma_1 - \sigma_1^4)\sigma_2^5 + (\gamma_2 - \sigma_2^4)\sigma_1^5}{3(\sigma_1\sigma_2(\sigma_1 + \sigma_2))^3}.$$

For  $D_A$ -optimal biased coin design,

$$(3.2) \quad \sqrt{l} \left( \frac{n_{l,1}}{l} - \frac{\sigma_1^{4/3}}{\sigma_1^{4/3} + \sigma_2^{4/3}} \right) \rightarrow N(0, \eta_1^2),$$

where

$$\eta_1^2 = \frac{\sigma_1^{4/3}\sigma_2^{4/3}}{5(\sigma_1^{4/3} + \sigma_2^{4/3})^2} + \frac{8(\gamma_1 - \sigma_1^4)\sigma_2^{16/3} + (\gamma_2 - \sigma_2^4)\sigma_1^{16/3}}{15\sigma_1^{8/3}\sigma_2^{8/3}(\sigma_1^{4/3} + \sigma_2^{4/3})^3}.$$

*Proof.* We just have to check Conditions 1 and 2. Now  $\boldsymbol{\theta} = (\mu_1, \sigma_1^2, \mu_2, \sigma_2^2)'$ . Condition 2 is satisfied. To check Condition 1, we rewrite the parameter estimators as

$$\hat{\mu}_{l,i} = \bar{Y}_{l,i} = n_{l,i}^{-1} \sum_{j=1}^l X_{j,i} Y_j$$

and

$$\hat{\sigma}_{l,i}^2 = (n_{l,i} - 1)^{-1} \sum_{j=1}^l X_{j,i} (Y_j - \bar{Y}_{l,i})^2.$$

Further  $\hat{\sigma}_{l,i}^2$  can be rewritten as

$$\hat{\sigma}_{l,i}^2 = (n_{l,i})^{-1} \sum_{j=1}^l X_{j,i} (Y_j - \mu_i)^2 + o(n_{l,i}^{-1/2}).$$

Therefore,

$$\hat{\mu}_{l,i} - \mu_i = n_{l,i}^{-1} \sum_{j=1}^l X_{j,i} (Y_j - \mu_i)$$

and

$$\hat{\sigma}_{l,i}^2 - \sigma_i^2 = (n_{l,i})^{-1} \sum_{j=1}^l X_{j,i} [(Y_j - \mu_i)^2 - \sigma_i^2] + o(n_{l,i}^{-1/2}).$$

Condition 1 is also satisfied.

From Condition 1, we can calculate  $\mathbf{V}_i$  ( $i = 1, 2$ ) as

$$\mathbf{V}_i = E \mathbf{h}_i(Y_j, \boldsymbol{\theta}) (\mathbf{h}_i(Y_j, \boldsymbol{\theta}))' = \begin{pmatrix} \sigma_i^2 & \xi_i \\ \xi_i & \gamma_i \end{pmatrix},$$

where  $\xi_1 = E(e_j^3 | \mathbf{X}_j = (1, 0)')$  and  $\xi_2 = E(e_j^3 | \mathbf{X}_j = (0, 1)')$ . By some routine calculations, we obtain Corollary 3.  $\square$

**Remark 3.1.** When  $e_j$  are normally distributed, then, we have the following results. Note that  $\gamma_1 = 3\sigma_1^4$  and  $\gamma_2 = 3\sigma_2^4$ . Therefore,

$$(3.3) \quad \sqrt{l} \left( \frac{n_{l,1}}{l} - \frac{\sigma_1}{\sigma_1 + \sigma_2} \right) \rightarrow N \left( \frac{\sigma_1}{\sigma_1 + \sigma_2}, \frac{\sigma_1\sigma_2}{(\sigma_1 + \sigma_2)^2} \right).$$

When  $D_A$ -optimal biased coin design is used, then

$$(3.4) \quad \sqrt{l} \left( \frac{n_{l,1}}{l} - \frac{\sigma_1^{4/3}}{\sigma_1^{4/3} + \sigma_2^{4/3}} \right) \rightarrow N \left( 0, \frac{19(\sigma_1\sigma_2)^{4/3}}{15(\sigma_1^{4/3} + \sigma_2^{4/3})^2} \right).$$

**EXAMPLE 2 (DICHOTOMOUS OUTCOMES).** For a two-arm clinical trial with binary outcomes  $Y_{l,i}$ ,  $i = 1, 2$ , (i.e., success or failure) with success probabilities  $p_1$  and  $p_2$  respectively, the MLEs of the respective parameters can be inserted into  $P_{l+1,1}$ . In this case,  $\boldsymbol{\theta} = (p_1, p_2)'$ . We can then obtain the  $D$ -optimal and  $D_A$ -optimal biased coin designs. Now let

$$\begin{aligned} y_{l,i} &= \text{number of successes on treatment } i \text{ at stage } l \\ n_{l,i} &= \text{number of patients on treatment } i \text{ at stage } l \\ \hat{p}_{l,i} &= y_{l,i}/n_{l,i} \\ \hat{q}_{l,i} &= 1 - \hat{p}_{l,i}. \end{aligned}$$

For  $D$ -optimal biased coin design,  $\rho(\boldsymbol{\theta}) = \sqrt{p_1 q_1 / (\sqrt{p_1 q_1} + \sqrt{p_2 q_2})}$ , where  $q_i = 1 - p_i$  and  $\theta_i = p_i$ .

$$\begin{aligned} \rho(\boldsymbol{\theta}) &= \frac{\sigma_1}{\sigma_1 + \sigma_2} \\ &= \frac{\sqrt{E(Y_{l,1}^2) - E(Y_{l,1})^2}}{\sqrt{E(Y_{l,1}^2) - E(Y_{l,1})^2} + \sqrt{E(Y_{l,2}^2) - E(Y_{l,2})^2}} \\ &= \frac{(\theta_1 - \theta_1^2)^{1/2}}{(\theta_1 - \theta_1^2)^{1/2} + (\theta_2 - \theta_2^2)^{1/2}}. \end{aligned}$$

Condition 2 is satisfied. It is easy to see that Condition 1 is also satisfied. The gradient of  $\rho$  can be calculated as

$$(\nabla(\rho)|_{\boldsymbol{\theta}})' = (\nabla(\rho)|_{\theta_1}, \nabla(\rho)|_{\theta_2}),$$

with

$$\nabla(\rho)|_{\theta_1} = \frac{(1 - 2p_1)(p_2 - p_2^2)^{1/2}}{2(p_1 - p_1^2)^{1/2} ((p_1 - p_1^2)^{1/2} + (p_2 - p_2^2)^{1/2})^2}$$

and

$$\nabla(\rho)|_{\theta_2} = \frac{(1 - 2p_2)(p_1 - p_1^2)^{1/2}}{2(p_2 - p_2^2)^{1/2} ((p_1 - p_1^2)^{1/2} + (p_2 - p_2^2)^{1/2})^2}.$$

The matrix  $\mathbf{V}$  is

$$\mathbf{v} = \begin{pmatrix} (p_1 - p_1^2)^{\frac{1}{2}} ((p_1 - p_1^2)^{\frac{1}{2}} + (p_2 - p_2^2)^{\frac{1}{2}}) & 0 \\ 0 & (p_2 - p_2^2)^{\frac{1}{2}} ((p_1 - p_1^2)^{\frac{1}{2}} + (p_2 - p_2^2)^{\frac{1}{2}}) \end{pmatrix}.$$

Together, we get

$$\begin{aligned} \tau^2 &= \frac{(1 - 2p_1)^2 (p_2 - p_2^2)^{3/2} + (1 - 2p_2)^2 (p_1 - p_1^2)^{3/2}}{4(p_1 - p_1^2)^{1/2} (p_2 - p_2^2)^{1/2} ((p_1 - p_1^2)^{1/2} + (p_2 - p_2^2)^{1/2})^3} \\ &= \frac{(1 - 2p_1)^2 (p_2 q_2)^{3/2} + (1 - 2p_2)^2 (p_1 q_1)^{3/2}}{4(p_1 q_1)^{1/2} (p_2 q_2)^{1/2} ((p_1 q_1)^{1/2} + (p_2 q_2)^{1/2})^3}. \end{aligned}$$

Similarly, we can obtain the results for  $D_A$ -optimal biased coin design. Here we state the results.

**Corollary 4.** For  $D$ -optimal biased coin design,

$$(3.5) \quad \sqrt{l} \left( \frac{n_{l,1}}{l} - \frac{(p_1 q_1)^{1/2}}{(p_1 q_1)^{1/2} + (p_2 q_2)^{1/2}} \right) \rightarrow N(0, \eta_2^2),$$

where

$$\eta_2^2 = \frac{\sqrt{(p_1 q_1)(p_2 q_2)}}{3 \left( (p_1 q_1)^{1/2} + (p_2 q_2)^{1/2} \right)^2} + \frac{(1 - 2p_1)^2 (p_2 q_2)^{3/2} + (1 - 2p_2)^2 (p_1 q_1)^{3/2}}{3 (p_1 q_1)^{1/2} (p_2 q_2)^{1/2} \left( (p_1 q_1)^{1/2} + (p_2 q_2)^{1/2} \right)^3}.$$

For  $D_A$ -optimal biased coin design,

$$(3.6) \quad \sqrt{l} \left( \frac{n_{l,1}}{l} - \frac{(p_1 q_1)^{2/3}}{(p_1 q_1)^{2/3} + (p_2 q_2)^{2/3}} \right) \rightarrow N(0, \eta_3^2),$$

where

$$\eta_3^2 = \frac{1}{5} \left( \frac{(p_1 q_1)^{2/3} (p_2 q_2)^{2/3}}{\left( (p_1 q_1)^{2/3} + (p_2 q_2)^{2/3} \right)^2} \right) + \left( \frac{2 \left( (1 - 2p_1)^2 (p_2 q_2)^{5/3} + (1 - 2p_2)^2 (p_1 q_1)^{5/3} \right)}{15 (p_1 q_1)^{1/3} (p_2 q_2)^{1/3} \left( (p_1 q_1)^{2/3} + (p_2 q_2)^{2/3} \right)^3} \right).$$

## 4. SIMULATION STUDIES

In Section 2 and 3, we obtained some theoretical properties of the  $D$ -optimal and  $D_A$ -optimal biased coin designs. As studied in Hu and Rosenberger (2003), these properties play a very important role for randomized designs. However, we need to know whether the asymptotic results work

for small sample cases. Here we study the properties of the two allocation schemes when the outcome was normal-distributed or dichotomous through simulation studies. We also made comparisons to the randomized 50:50 allocation scheme in terms of power.

### 4.1 Allocation proportion and variability

First, we examined the performance of the  $D$ -optimal biased coin under Example 1 with normal distribution. We implemented 1000 simulated clinical trials, fixing  $\mu_1 = \mu_2 = 0$  and  $\sigma_1^2 = 1$ . We considered a wide range of values for other parameters:  $\sigma_2^2 = 1, 4, 9, 16$  and  $25$ ; the total sample size  $N = 50, 100, 200, 400$  and  $800$  (see Figures 1 and 2, in the columns from left to right); and the initial number of patients assigned to each treatment  $m = 5, 10$  and  $20$  (in the rows from the bottom up in Figures 1 and 2). Each panel of Figure 1 shows the absolute difference between the theoretical mean of allocation fractions and the empirical mean computed over the 1000 simulations versus a wide range of outcome variance values under treatment 2. All the possible combinations of  $N$  and  $m$  are exhibited. Figure 2 shows the corresponding absolute difference of the variances obtained from theoretical derivation and simulations.

Our findings showed that the empirical results were different from the theoretical mean when  $N$  was small and  $m$  was large. We believed the results are reasonable, i.e., if 20 patients were assigned to each treatment arm and  $N = 50$ , then only 10 patients were available to the  $D$ -optimal biased coin allocation procedure. So the sample size was too small for the asymptotic results to be correct. The greatest difference was observed when  $\sigma_2^2$  was the largest, because the number of patients assigned to the two treatments were expected to be the most different. Overall, it did show that

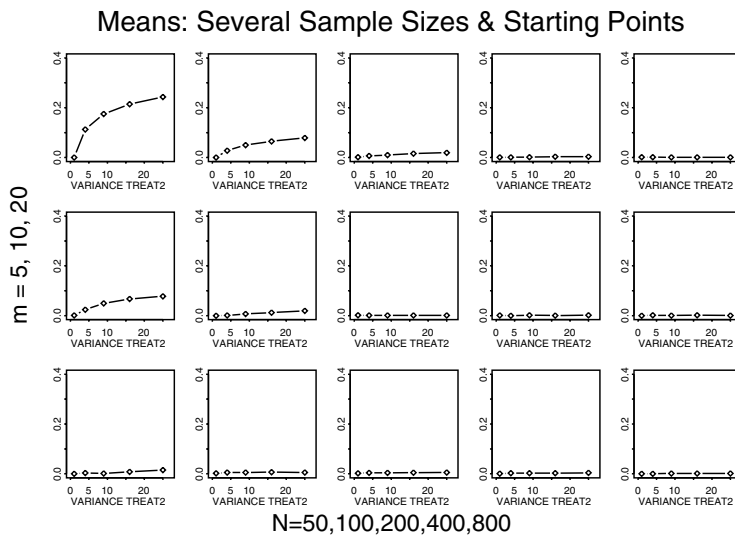


Figure 1. Comparisons Between Theoretical and Empirical Means of Allocation Proportions, Normal-Distributed Outcome and  $D$ -Optimal Biased Coin Design.

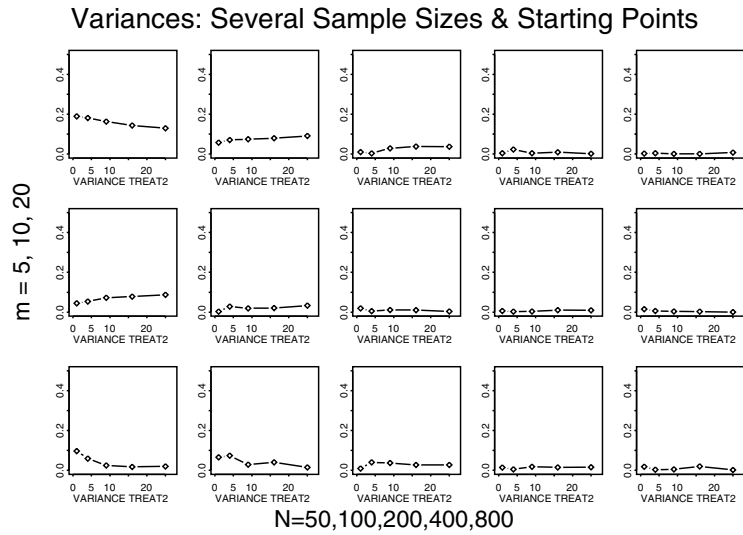


Figure 2. Comparisons Between Theoretical and Empirical Variances of Allocation Proportions, Normal-Distributed Outcome and D-Optimal Biased Coin Design.

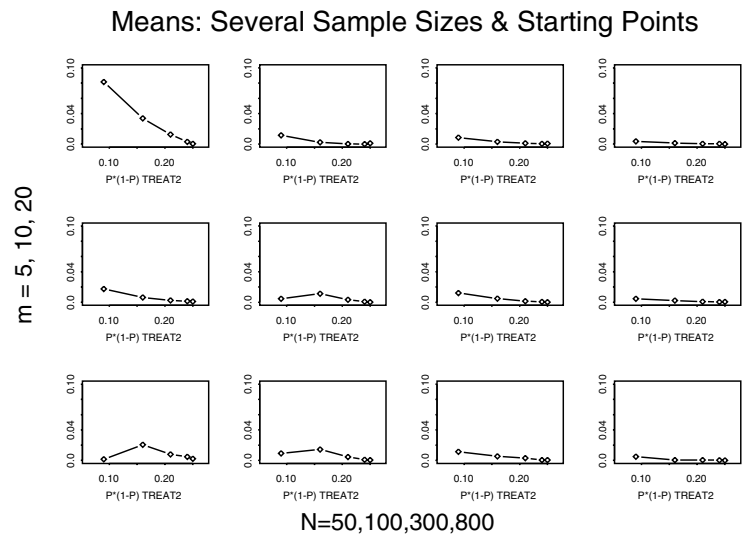


Figure 3. Comparisons Between Theoretical and Empirical Means of Allocation Proportions, Binary Outcome and D-optimal Biased Coin Design.

the expected mean allocation fraction and the empirical result converged fairly quickly as sample size increased. Moreover, we noticed that the smaller value of  $N$  induced a large amount of variability in the allocation fractions. This could be because of the great variability in estimating the experimental outcome variances with only a few data points in the allocation procedure. With the other parameters used, the sample variance of allocation fractions was fairly close to the theoretical results, in particular, the difference between the two decreased as sample size increased. These simulations supported the choice of  $m = 5$ .

Next, we studied the case of a dichotomous outcome as in Example 2. We fixed  $p_1(1 - p_1) = 0.25$  and considered a range of values for other parameters:  $p_2(1 - p_2) = 0.09, 0.16, 0.21, 0.24,$  and  $0.25$ ;  $m = 5, 10,$  and  $20$ ;  $N = 50, 100, 300,$  and  $800$ . In each scenario, we implemented 1000 simulations. Figure 3 shows the absolute difference between the theoretical mean of the allocation fraction and the empirical results obtained across the simulations. The results were similar to those of the normal distribution case. Large  $m$  and small  $N$  produced the largest discrepancy between the theoretical and simulated results. We ob-

### Variates: Several Sample Sizes & Starting Points

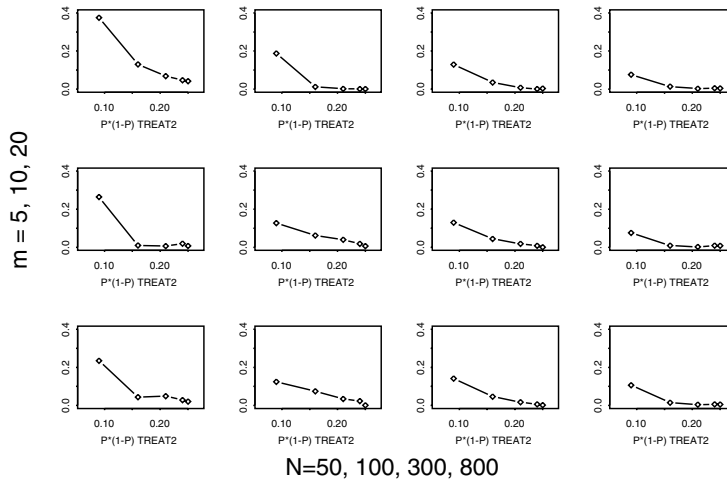


Figure 4. Comparisons Between Theoretical and Empirical Variances of Allocation Proportions, Binary Outcome and  $D$ -Optimal Biased Coin Design.

served that the most unbalanced design was expected when  $p_2(1 - p_2) = 0.09$ .

The variance results are exhibited in Figure 4. In all cases we noted that when the success probability under treatment 2 approached 1, a relatively larger deviation from the expected value was produced. We also observed that the discrepancy between the expected variance and the empirical sample variance tended toward 0 as  $N$  increased. These simulations also supported the choice of  $m = 5$ .

We have also investigated the accuracy of the theoretical results using the generalized  $D_A$ -optimal biased coin design via sets of simulation studies. We observed similar results as  $D$ -optimal biased coin design.

### 4.2 Power analysis

We implemented another set of simulations to study the power that the proposed allocation scheme could achieve, and also compared the results to those of the randomized 50:50 allocation. We simulated 4 sets of 5,000 clinical trials for the normally distributed outcome. In each simulated clinical trial, the one-sided  $t$ -test controlling for a type-I error rate of 0.05 was considered. The total sample size required to detect a mean difference of 0.5 with the power of 80% was calculated in the case of a randomized 50:50 allocation. We then used this sample size as  $N$  in the generalized  $D$ -optimal (or  $D_A$ -optimal) biased coin design with  $m = 10$ . The data was simulated as follows:

- Treatment 1:  $Y_1 \sim N(1, 1)$ , and
- Treatment 2:  $Y_2 \sim N(1.5, \sigma_2^2)$ ,  $\sigma_2 = 1, \dots, 4$ .

For each set of postulated distributions, allocations preceding the subsequent tests were made, respectively, using the  $D$ -optimal (or  $D_A$ -optimal) biased coins and the balanced procedure. We report the proportion of times that

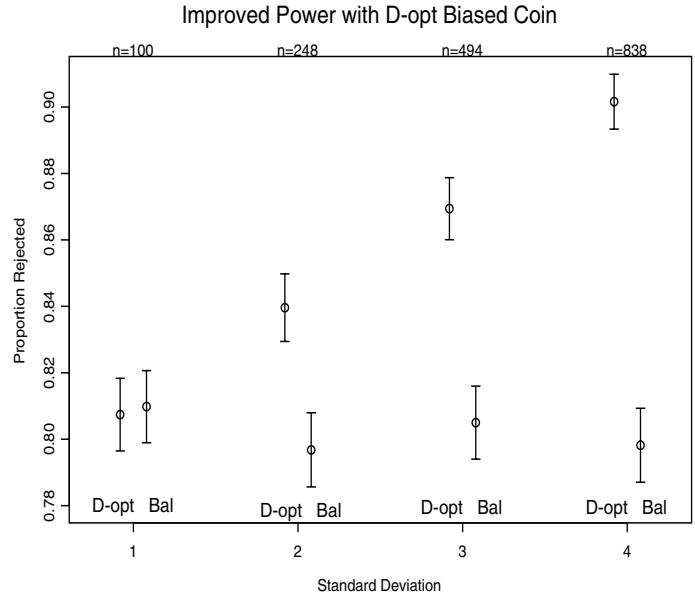


Figure 5. Power Comparisons Between  $D$ -Optimal Biased Coin Design and Randomized 50:50 Allocation.

$H_0 (\mu_1 = \mu_2)$  was rejected in Figure 5 ( $D$ -optimal) and Figure 6 ( $D_A$ -optimal), together with the 95% CI. The results indicate that using both  $D$ -optimal and  $D_A$ -optimal biased coins provided larger detecting power than the balanced allocation. The advantage became substantial as the difference of the outcome variances under different treatments increased. As a result, fewer subjects could be recruited to achieve the same desired test aim, thus reducing the cost or, potentially, the patient risk to some extent.



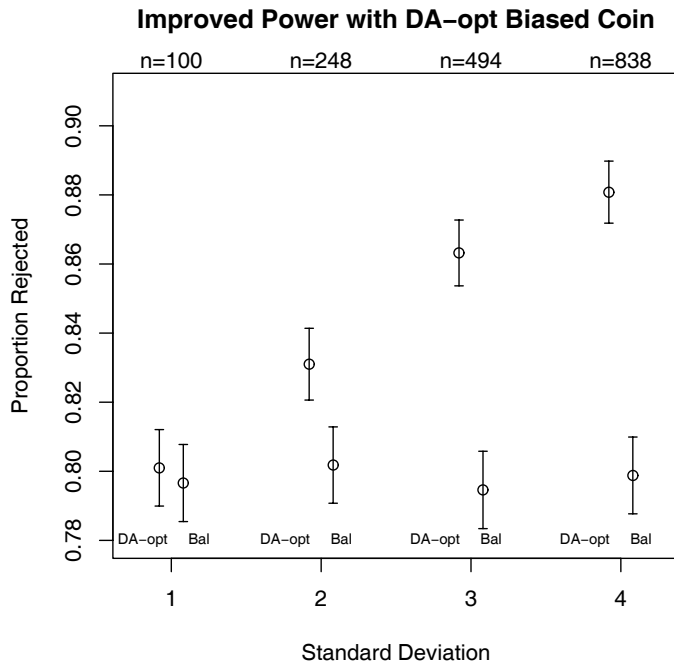


Figure 6. Power Comparisons Between  $D_A$ -Optimal Biased Coin Design and Randomized 50:50 Allocation.

## 5. DISCUSSION

We have proposed generalized  $D$ -optimal and  $D_A$ -optimal response adaptive biased coin designs without any strong distributional assumptions on the treatment outcomes. We have shown that the allocation proportions of the generalized methods have asymptotically normal distribution properties. We have also observed that Atkinson's design with known equal variances of responses under two treatments is a special case of the generalized design, therefore also following asymptotically normal distribution with mean  $1/2$ , a balanced design. The asymptotic distribution properties can be useful in practice in at least two ways. One possible application is for a researcher to obtain an assessment of the possible patient allocation pattern and variation of that allocation before deciding to use a randomization scheme. In addition, one can also use these asymptotic results with the tools provided in Hu and Rosenberger (2003) to compare the design scheme to other adaptive designs with respect to the power of the tests. When the variances are unknown, the proposed design is response-adaptive design. Some implementation issues are discussed in Rosenberger and Lachin (2002).

In real clinical trials, there is no logistical difficulty in incorporating delayed responses into the generalized  $D$ -optimal and  $D_A$ -optimal response adaptive biased coin designs, provided some responses become available during the recruitment and randomization period. The unknown variances can be estimated when data become available. By using similar technical as Bai, Hu, and Rosenberger (2002), we can

show that the asymptotic results of the previous sections will hold under the following widely satisfied condition: the probability that at least  $m$  additional patients will arrive prior to a patient's response is of order  $o(m^{-c})$  for some positive  $c$ .

We have been focused on the large sample properties of the generalized  $D$ -optimal and  $D_A$ -optimal response adaptive biased coin designs. In application, it is important to know the suitable size of a clinical trial that is sufficiently large that the asymptotic properties are valid. Based on the simulations in Section 4, we found that the simulated and asymptotic proportions (and variances) agreed quite well for sample size  $n = 100$ . More simulation results can be found in Gwise (2005). Hu and Rosenberger (2003) and Rosenberger and Hu (2004) investigated the accuracy of the large approximations for moderate sample sizes using simulation for other response-adaptive designs. For example, the doubly adaptive biased coin design (Hu and Zhang, 2004) with  $\gamma = 2$  attained the target allocation quite accurately, and the simulated and asymptotic variances agreed quite well for sample size  $n = 100$ .

Real clinical trials often evaluate more than two treatments or drugs at one time. It is thus worthwhile to pursue solutions for more complicated cases. In addition, other patient information (e.g., smoking status), which might directly affect the treatment outcome might be available for consideration in the trial design. Utilizing important clinical information to increase the testing power is a promising future research topic.

## ACKNOWLEDGEMENTS

This work was supported by NSF Grant DMS-0349048.

Special thanks go to an anonymous referee for the constructive comments, which led to an improved version of the paper.

## APPENDIX: PROOFS

The asymptotic properties of the urn models and sequential estimation based designs are well studied in the literature. Bai and Hu (2005) studied asymptotic normality for general randomized urn model. Hu and Zhang (2004) obtained some important asymptotic properties for doubly adaptive biased-coin designs. The idea of the  $D$ -optimal biased coin design in Section 2 is different from doubly adaptive biased-coin design. The sequential estimation based designs usually have a predetermined target allocation function, which depends on some sequential estimates of some unknown parameters. The  $D$ -optimal adaptive biased-coin design also depends on the sequential estimates of the unknown variances. However, it does not target any predetermined allocation proportion. The proof of Theorem 1 includes two steps: (i) First we try to find an artificial target proportion to make the  $D$ -optimal biased coin design a special case of doubly adaptive biased-coin design; (ii) then

we use the result from Hu and Zhang (2004) to obtain the asymptotic properties of the  $D$ -optimal biased coin design by verifying all the conditions.

*Proof of Theorem 1.* First we define

$$(A.1) \quad w_l = \frac{n_{l,1}}{l}, \quad u = \frac{\sigma_1}{\sigma_1 + \sigma_2} \quad \text{and} \quad \hat{u}_l = \frac{\hat{\sigma}_{l,1}}{\hat{\sigma}_{l,1} + \hat{\sigma}_{l,2}},$$

where  $u$  is the target allocation,  $\hat{u}_l$  is the estimated target allocation after  $l$  observations and  $w_l$  is the actual allocation proportion after  $l$  patients being assigned. Then expression (2.6) ( $D$ -optimal biased coin design) can be rewritten as

$$(A.2) \quad P_{l+1,1} = \frac{\sigma_1^2/n_{l,1}}{\sigma_1^2/n_{l,1} + \sigma_2^2/n_{l,2}} = \frac{u(\frac{u}{w})}{u(\frac{u}{w}) + (1-u)(\frac{1-u}{1-w})}.$$

Therefore, we can view the  $D$ -optimal biased-coin design as a doubly adaptive biased-coin design with the allocation function:

$$(A.3) \quad g(0, y) = 1, \quad g(1, y) = 0, \quad g(x, y) = \frac{y(\frac{y}{x})}{y(\frac{y}{x}) + (1-y)(\frac{1-y}{1-x})}.$$

Thus, we can use Theorem 2.1 of Hu and Zhang (2004) by verifying all the conditions. It is easy to see that the conditions (i)–(iii), (vi) and (vii) are satisfied. However, Hu and Zhang (2004) assumes that the parameter  $\theta$  is the expectation of the responses. In some applications, this can be achieved by some transformations. But in some other applications, this may not be satisfied. Here, we consider a more general estimation of  $\theta$  as in Condition 1. Therefore, we need to check the proofs of Hu and Zhang (2004) under Condition 1 and 2.

In the proofs of Hu and Zhang (2004), the Lemma A.1, Lemma A.2 and Lemma A.3 hold, because they do not involve  $\theta$  and  $\hat{\theta}_l$ . Now we show Lemma A.4 from Condition 1. Based on Condition 1, we have

$$\hat{\theta}_{l,i} - \theta_i = \frac{1}{n_{l,i}} \sum_{j=1}^{n_{l,i}} \mathbf{h}_i(Y_{j,i}, \theta) + o(n_{l,i}^{-1/2}),$$

Here  $Y_{j,i}$  are the observations corresponding the treatment  $i$ . Because  $E\|\mathbf{h}_i(Y_{j,i}, \theta)\|^2 < \infty$ , we obtain Lemma A.4 by applying the law of iterated logarithm to  $\sum_{j=1}^{n_{l,i}} \mathbf{h}_i(Y_{j,i}, \theta)$ . Lemma A.5 is a direct result from Condition 2 and  $\rho(\theta) = \sigma_1(\theta)/(\sigma_1(\theta) + \sigma_2(\theta)) > 0$ .

Based on the proof of Theorem 4.1 of Hu and Zhang (2004), we can show the strong consistency of  $n_{l,1}$  (result (i) of Theorem 1) similarly. To show the law of iterated logarithm (result (ii) of Theorem 1), we can use similar proof of Theorem 4.2 of Hu and Zhang (2004) by noting the Condition 1 and Condition 2. From Condition 1, we have

$$\hat{\theta}_l - \theta = O\left(\sqrt{\frac{\log \log(l)}{l}}\right) \text{ almost surely}$$

and then, by Condition 2,

$$\begin{aligned} \rho(\hat{\theta}_l) - \rho(\theta) &= \frac{\sigma_1(\hat{\theta}_l)}{\sigma_1(\hat{\theta}_l) + \sigma_2(\hat{\theta}_l)} - \frac{\sigma_1}{\sigma_1 + \sigma_2} \\ &= O\left(\sqrt{\frac{\log \log(l)}{l}}\right) \text{ almost surely.} \end{aligned}$$

The result (ii) is then followed by the law of iterated logarithm of martingale.

In the proof of Theorem 4.3 in Hu and Zhang (2004), because  $\theta$  can be expressed as the expectation of the random vector  $\mathbf{Y}$  (or some transformation of  $\mathbf{Y}$ ), then the estimator  $\hat{\theta}$  can be represented as a certain type of sample average. Here we can not use Theorem 4.3 of Hu and Zhang (2004) directly. However, we can do the following modification. Let  $v_1 = \sigma_1(\theta)/(\sigma_1(\theta) + \sigma_2(\theta))$  and  $v_2 = 1 - v_1$ . We define

$$\mathbf{Q}_l = \sum_{j=1}^l \Delta \mathbf{Q}_j, \quad \text{where } \Delta \mathbf{Q}_j = (\Delta \mathbf{Q}'_{j,1}, \Delta \mathbf{Q}'_{j,2})'$$

and

$$\Delta \mathbf{Q}_{j,i} = X_{j,i} \mathbf{h}_i(Y_j, \theta)/v_i.$$

Based on Condition 1, we have

$$(A.4) \quad \hat{\theta}_l - \theta = \frac{\mathbf{Q}_l}{l} + O\left(\frac{\log \log l}{l}\right) \text{ almost surely.}$$

From Condition 2, we have

$$(A.5) \quad \begin{aligned} \rho(\hat{\theta}_l) - \rho(\theta) &= (\hat{\theta}_l - \theta) \nabla(\rho)|_{\theta} + o\left(\|\hat{\theta}_l - \theta\|^{1+\delta}\right) \\ &= (\hat{\theta}_l - \theta) \nabla(\rho)|_{\theta} + o(m^{-1/2-\delta/2}), \end{aligned}$$

where  $\delta$  is a positive constant. Thus, we obtain the new version of the equations (A.10) and (A.11) of Hu and Zhang (2004). Based on (A.5), we can then obtain similar expression of (A.12) for  $n_{l,1} - l\rho(\theta)$ . We can then obtain the asymptotic normality of  $n_{l,1}$  similarly as Theorem 4.3 of Hu and Zhang (2004). The technique details are not included here. The asymptotic variance can be calculated by using (A.4), (A.5) and the modified  $\Delta \mathbf{Q}_j$ .

Now we use the result of Theorem 2.1 of Hu and Zhang (2004) and (A.5) to calculate the asymptotic variance in Theorem 1. First, it is easy to see that

$$\lambda = \frac{\partial g}{\partial x} \Big|_{(\rho, \rho)} = -1 \quad \text{and} \quad \gamma = \frac{\partial g}{\partial y} \Big|_{(\rho, \rho)} = 2.$$

Therefore, the variance is then

$$\frac{1}{3} \frac{\sigma_1 \sigma_2}{(\sigma_1 + \sigma_2)^2} + \frac{4}{3} (\nabla(\rho)|_{\theta})' \mathbf{V}(\nabla(\rho)|_{\theta}).$$

□

*Proof of Theorem 2.* For  $D_A$ -optimal biased coin design, we define

$$(A.6) \quad w_l = \frac{n_{l,1}}{l}, \quad u = \frac{\sigma_1^{4/3}}{\sigma_1^{4/3} + \sigma_2^{4/3}} \quad \text{and} \quad \hat{u}_l = \frac{\hat{\sigma}_{l,1}^{4/3}}{\hat{\sigma}_{l,1}^{4/3} + \hat{\sigma}_{l,2}^{4/3}},$$

where  $u$  is the target allocation,  $\hat{u}_l$  is the estimated target allocation after  $l$  observations and  $w_l$  is the actual allocation proportion after  $l$  patients being assigned. Then the allocation probability can be rewritten as

$$(A.7) \quad P_{l+1,1} = \frac{(\sigma_1^2/n_{l,1})^2}{(\sigma_1^2/n_{l,1})^2 + (\sigma_2^2/n_{l,2})^2} = \frac{u(\frac{u}{w})^2}{u(\frac{u}{w})^2 + (1-u)(\frac{1-u}{1-w})^2}.$$

Therefore, we can view the  $D$ -optimal biased-coin design as a doubly adaptive biased-coin design with the allocation function:

$$(A.8) \quad g(0, y) = 1, \quad g(1, y) = 0, \\ g(x, y) = \frac{y(\frac{y}{x})^2}{y(\frac{y}{x})^2 + (1-y)(\frac{1-y}{1-x})^2}.$$

The rest is the same as the proof of Theorem 1.  $\square$

Received 10 November 2007

## REFERENCES

- Athreya, K. B. and Karlin, S. (1968). Embedding of urn schemes into continuous time Markov branching processes and related limit theorems. *The Annals of Mathematical Statistics* **39** 1801–1817. [MR0232455](#)
- Atkinson, A. C. (1982). Optimum biased coin designs for sequential clinical trials with prognostic factors. *Biometrika* **69** 61–67. [MR0655670](#)
- Bai, Z. D. and Hu, F. (2005). Asymptotics in randomized urn models. *Annals of Applied Probability* **15** 914–940. [MR2114994](#)
- Bai, Z. D., Hu, F. and Rosenberger, W. F. (2002). Asymptotic properties of adaptive designs for clinical trials with delayed response. *Annals of Statistics* **30** 122–139. [MR1892658](#)
- Efron, B. (1971). Forcing a sequential experiment to be balanced. *Biometrika* **58** 403–417. [MR0312660](#)
- Eisele, J. R. (1994). The doubly adaptive biased coin design for sequential clinical trials. *Journal of Statistical Planning and Inference* **38** 249–262. [MR1256599](#)

- Eisele, J. R. and Woodroffe, M. B. (1995). Central limit theorems for doubly adaptive biased coin designs. *The Annals of Statistics* **1** 234–254. [MR1331666](#)
- Gwise, T. E. (2005). *Optimal biased coin designs for clinical trials*, Ph.D. thesis, Department of Statistics, University of Virginia.
- Hu, F. and Rosenberger, W. F. (2003). Evaluating response-adaptive randomization procedures for treatment comparisons. *Journal of the American Statistical Association* **98** 671–678. [MR2011680](#)
- Hu, F. and Rosenberger, W. F. (2006). *The Theory of Response-Adaptive Randomization in Clinical Trials*, John Wiley and Sons, Inc., New York. [MR2245329](#)
- Hu, F. and Zhang, L. (2004). Asymptotic properties of doubly adaptive biased coin designs for multitreatment clinical trials. *The Annals of Statistics* **32** 268–301. [MR2051008](#)
- Jennison, C. and Turnbull, B. W. (2000). *Group Sequential Methods with Applications to Clinical Trials*. Chapman and Hall/CRC Press, Boca Raton, FL. [MR1710781](#)
- Keifer, J. and Wolfowitz, J. (1960). The equivalence of two extremum problems. *Canadian Journal of Mathematics* **12** 363–366. [MR0117842](#)
- Rosenberger, W. F. (2002). Randomized urn models and sequential design. *Sequential Analysis* **21** 1–28. [MR1903097](#)
- Rosenberger, W. F. and Hu, F. (2004). Maximizing power and minimizing treatment failures in clinical trials. *Clinical Trials* **1** 141–147.
- Rosenberger, W. F. and Lachin, J. M. (2002). *Randomization in Clinical Trials Theory and Practice*, John Wiley and Sons, Inc., New York. [MR1914364](#)
- Tymofeyev, Y., Rosenberger, W. F., and Hu, F. (2007). Implementing optimal allocation in sequential binary response experiments. *Journal of the American Statistical Association* **102** 224–234. [MR2345540](#)
- Wei, L. J. (1977). A class of designs for sequential clinical trials. *Journal of the American Statistical Association* **72** 382–386.
- Wei, L. J. (1978). The adaptive biased coin design for sequential experiments. *The Annals of Statistics* **6** 92–100. [MR0471205](#)
- Wei, L. J. and Durham, S. (1978). The randomized play the winner rule in medical trials. *Journal of the American Statistical Association* **73** 840–843.

Thomas E. Gwise  
Food and Drug Administration, USA

Jianhua Hu  
Anderson Cancer Center, USA

Feifang Hu  
Department of Statistics  
University of Virginia  
Charlottesville  
Virginia 22904-4135, USA  
E-mail address: [fh6e@virginia.edu](mailto:fh6e@virginia.edu)