Maximizing power and minimizing treatment failures in clinical trials

William F Rosenberger and Feifang Hu

Clin Trials 2004 1: 141
DOI: 10.1191/1740774504cn016oa

CLINICAL TRIALS ARTICLE Clinical Trials 2004; 1: 141–147

Maximizing power and minimizing treatment failures in clinical trials

William F Rosenberger[a,b] and Feifang Hu[c,d]

Background Response-adaptive randomization procedures have a long history in the theoretical statistics literature over the past four decades. The main idea historically was to develop randomization procedures that place fewer patients on the inferior treatment. More recent research has changed the main focus to that of usual considerations in typical clinical trials: power, sample size, expected treatment failures, maintaining randomization, among others.

Methods We describe response-adaptive randomization procedures for simple clinical trials comparing two binomial success probabilities, including the randomized play-the-winner rule, the drop-the-loser rule, and a modification of the doubly-adaptive biased coin design. We treat as our principal goal minimizing expected treatment failures while preserving power and randomization. Based on some recent theoretical literature [1,8,14], the basic guidelines for selecting an appropriate procedure include targeting optimal allocation, having small variability, and preserving randomization. We use simulation to compare power and expected treatment failures according to these guidelines.

Results When the two treatments had high probabilities (>0.5) of success, the randomized play-the-winner rule was less powerful than complete randomization and the drop-the-loser rule by 1–3 percent with slightly larger expected number of treatment failures than the drop-the-loser rule. For all the success probabilities we examined, the drop-the-loser rule was within 1 percent of the power of complete randomization with a modest reduction of treatment failures. The doubly-adaptive biased coin design was as powerful or slightly more powerful than complete randomization in every case and expected treatment failures were always less, with modest reductions of the order of 0.3 percent to 8.3 percent.

Conclusions We conclude that the drop-the-loser rule and a modification of the doubly-adaptive biased coin design are the preferred procedures, and simulations show that these procedures yield a modest reduction in expected treatment failures while preserving power over complete randomization. *Clinical Trials* 2004; **1**: 141–147. www.SCTjournal.com

Introduction

Consider the simplest clinical trial of two treatments with a binary outcome. The probability of a success on treatment A is denoted P_A and the probability of a success on treatment B is denoted P_B, with $Q_A = 1 - P_A$ and $Q_B = 1 - P_B$. Let n_A be the number of patients to be assigned to treatment A and n_B be the number of patients to be assigned to treatment B, so that $n_A + n_B = n$, where n is the total sample size. Let us assume that the usual Z-test for the difference of proportions is used, i.e.,

$$Z = \frac{\hat{P}_A - \hat{P}_B}{\sqrt{n\bar{P}\bar{Q}/n_A n_B}} \quad (1)$$

[a]Department of Mathematics and Statistics, University of Maryland, Baltimore County, Baltimore, MD, USA, [b]Department of Epidemiology and Preventive Medicine, University of Maryland School of Medicine, Baltimore, MD, USA, [c]Department of Statistics, University of Virginia, Charlottesville, VA, USA, [d]Division of Biostatistics and Epidemiology, Department of Health Evaluation Sciences, University of Virginia School of Medicine, Charlottesville, VA, USA
Author for correspondence: William F. Rosenberger, Department of Mathematics and Statistics, University of Maryland, Baltimore County, 1000 Hilltop Circle, Baltimore, MD 21250. E-mail: billr@math.umbc.edu

10.1191/1740774504cn016oa

where $\bar{P} = (\hat{P}_A + \hat{P}_B)/2$, $\bar{Q} = 1 - \bar{P}$, and $\hat{P}$ refers to the estimator.

The statistician, charged with computing an appropriate sample size, has a number of options. Traditionally, he would set $n_A = n_B = n/2$, fix the power of the Z-test, and find the minimum n to achieve that power. But there are alternate strategies if he does not insist on equal allocation. First he could fix the power, and find n_A and n_B to minimize the total sample size n. Secondly, he could fix the total sample size n and find n_A and n_B to maximize the power. It turns out that these two approaches have the same answer, as a little calculus will show:

$$n_A = \frac{n\sqrt{P_A Q_A}}{\sqrt{P_A Q_A} + \sqrt{P_B Q_B}}, \quad n_B = n - n_A \tag{2}$$

This is called *Neyman allocation*. Unfortunately, the statistician now has two dilemmas: 1) he doesn't know what P_A and P_B are; and 2) even if he did know, when $P_A + P_B > 1$, Neyman allocation will assign more patients to the less successful treatment, thereby introducing an ethical problem.

We will deal with dilemma (1) momentarily. Let us focus on the ethical dilemma (2). Based on the above results, a reasonable question that should interest the statistician is *what allocation will simultaneously maximize power and minimize the expected number of treatment failures?* Unfortunately, this question has no mathematical solution. However, the statistician can modify his approach to sample size computation as follows. He can fix power, and find n_A and n_B to minimize the expected number of treatment failures in the trial. Equivalently, he could fix the expected number of treatment failures and find n_A and n_B to maximize power. This leads to the following allocation:

$$n_A = \frac{n\sqrt{P_A}}{\sqrt{P_A} + \sqrt{P_B}}, \quad n_B = n - n_A \tag{3}$$

We will call this *optimal allocation* [1].

Now let us return to dilemma (1). Again, optimal allocation depends on the unknown probabilities P_A and P_B. Some authors (e.g., [2], p.41) argue that the statistician, at this point, has no alternative than to impose equal allocation to the two treatment groups, i.e., $n_A = n_B = n/2$. Formally, the argument is as follows: we begin the clinical trial operating under the null hypothesis, that $P_A = P_B$, and therefore both Neyman and optimal allocation reduce to equal allocation under that assumption. In [2], it is argued that equipoise, the state of genuine uncertainty about which treatment is better at the beginning of the trial, dictates that equal allocation always be used. Other authors have argued that, as data accrue in the sequential clinical trial, the statistician can begin to estimate P_A and P_B, and we can use these values to update the probability of randomization to A or B. This is the principle behind *response-adaptive randomization*.

In this paper, we describe a number of response-adaptive randomization procedures and evaluate them in terms of preserving power and minimizing expected treatment failures. We will have to deal with the trade-offs of power, variability and the degree of randomization. In so doing, we intend to present the best current procedure for use in actual clinical trials.

We consider only randomized, myopic procedures. Others have dealt with the problem from the far more difficult vantage of the two-armed bandit problem [3,4]. Most of the bandit literature deals with nonrandomized procedures and a globally optimal solution requires parallel processing. While our myopic procedures may not be globally optimal, our goal in this paper is to show that some of these procedures can be modestly attractive in terms of yielding smaller expected numbers of treatment failures with no loss of power. These procedures are simple to implement, as they requires only minor modification of the randomization code.

Response-adaptive randomization procedures that do not target optimal allocation

Response-adaptive procedures have been proposed in various forms over the past 40 years. Historically, they were proposed with the idea of developing a nonrandomized procedure that would eventually assign all patients to the better treatment. Such procedures do not fit well in the culture of modern clinical trials, where the predominant aim is to make decisions based on an appropriately powered, completely randomized, clinical trial.

Along another track in research on response-adaptive procedures, we have a sequence of procedures developed based on the idea that minimizing the expected failure rate is of most importance in clinical trials. This research led to randomized procedures using urn models, but these procedures were not developed in the context of preserving power. For most of these urn procedures, it can be shown that N_A/N_B (noting that N_A and N_B are now random variables, so are suitably capitalized) tends to the relative risk of failure Q_B/Q_B. Thus the proportion of patients assigned to treatment A in the limit will be $nQ_B/(Q_A + Q_B)$. We refer to this allocation as *urn allocation*, because it arises from randomized urn models. We refer to response-adaptive randomization procedures that result in urn allocation in the limit as procedures *targeting* urn allocation. Urn allocation has no formal

optimal properties. We now describe and compare some of these procedures.

Randomized play-the-winner rule

Wei and Durham's famous *randomized play-the-winner rule* [5] also uses an urn model to allocate treatments. This starts with a fixed number of type A balls and type B balls in an urn (computer generated). To randomize a patient, a ball is drawn, the corresponding treatment assigned, and the ball replaced. An additional ball of the same type is added if the patient's response is success and an additional ball of the opposite type is added if the patient's response is a failure. The randomized play-the-winner rule also targets urn allocation, but is fully randomized.

Drop-the-loser rule

Ivanova describes the following procedure [6]: an urn contains balls of three types, type A, type B, and type 0. When a patient is ready to be randomized, a ball is drawn at random. If it is type A or type B, the corresponding treatment is assigned and the patient's response is observed. If it is a success, the ball is replaced and the urn remains unchanged. If it is a failure, the ball is not replaced. If a type 0 ball is drawn, no subject is treated, and the ball is returned to the urn together with one ball of type A and one ball of type B. The type 0 balls ensure that the urn never gets depleted.

We will illustrate the procedure with a simple example. Suppose we start the trial with 1 ball of each type in the urn. The first patient is ready to be randomized, and a type A ball is drawn (with probability 1/3). The patient is a success, and the ball is replaced. The probability of assignment to A is still 1/3. Suppose for patient 2, a type 0 ball is drawn. That ball is returned to the urn, along with an additional type A and type B ball. Another ball is drawn. This time the probability that patient 2 will be assigned to A is 2/5 and to B is 2/5. Suppose a type A ball is drawn and we have a treatment failure. The ball is not returned to the urn, so now the probability that patient 3 will be assigned to A is now 1/4 and to B is 1/2.

The drop-the-loser rule can only target urn allocation as well. So we have two procedures that target the same allocation. The question arises, how bad is urn allocation in terms of preserving power? The answer largely stems from how far away it is from Neyman allocation. The answer is presented in Table 1. For larger values of P_A or P_B we see that it is very far from Neyman allocation and also very far from optimal allocation. Optimal allocation tends to split the difference between urn allocation and equal allocation, but unlike Neyman allocation, it always assigns more patients to the better treatment. When P_A and P_B are smaller (less than 0.5), urn allocation is much closer to Neyman allocation, and one can presume the procedures targeting urn allocation would be more powerful.

Table 1 Comparison of equal, Neyman, optimal, and urn allocation, proportion of patients assigned to treatment A

P_A	P_B	Equal	Neyman	Optimal	Urn
0.9	0.1	0.50	0.50	0.75	0.90
0.9	0.3	0.50	0.40[a]	0.63	0.88
0.9	0.5	0.50	0.38[a]	0.57	0.83
0.9	0.7	0.50	0.40[a]	0.53	0.75
0.9	0.8	0.50	0.43[a]	0.51	0.67
0.7	0.3	0.50	0.50	0.60	0.70
0.7	0.5	0.50	0.48[a]	0.54	0.63
0.5	0.4	0.50	0.51	0.53	0.55
0.3	0.1	0.50	0.60	0.63	0.56
0.2	0.1	0.50	0.57	0.59	0.53

[a]Allocates more patient to the inferior treatment.

The distribution of the usual Z-test, described in the Introduction, is asymptotically standard normal after using either the randomized play-the-winner or drop-the-loser rules [6,7]. Simulation results showing how close $E(N_A/n)$ is to urn allocation for moderate sample sizes have been reported elsewhere [8].

From [8], we can begin to formulate some guiding principles for choosing a response-adaptive randomization procedure. First, it should yield a limiting allocation close to optimal allocation. Second, a procedure with smaller variability in the allocation proportions will be more powerful. Third, a procedure should be randomized, to protect from the biases of nonrandomized studies [9]. We will examine these three procedures with respect to these criteria. First, none of the procedures yields a limiting allocation close to optimal allocation for large values of P_A or P_B. For the second criterion, it has been shown [6] that the drop-the-loser rule has smaller variability than the randomized play-the-winner rule.

Table 2 demonstrates the simulated power of the usual test statistic

$$Z = \frac{\hat{P}_A - \hat{P}_B}{\sqrt{n\bar{P}\bar{Q}/N_A N_B}},$$

where $\bar{P} = (\hat{P}_A + \hat{P}_B)/2$, $\bar{Q} = 1 - \bar{P}$, with $\alpha = 0.05$ (two-sided), for the three procedures. Also given is the simulated expected treatment failures and the standard deviation. The randomized play-the-winner rule was started with five balls of each type in the urn. The drop-the-loser rule had five type A balls, five type B balls, and one type 0 ball. The choice of initial urn composition has been somewhat controversial [10]. At one extreme, a small number

Table 2 Simulated power and expected treatment failures (standard deviation) for complete randomization and two response-adaptive randomization procedures

			Complete		RPW Rule		DL Rule	
P_A	P_B	n	Power	Failures	Power	Failures	Power	Failures
0.9	0.3	24	90	10 (2.4)	87	7 (2.4)	90	7 (1.8)
0.9	0.5	50	90	15 (3.2)	87	12 (3.2)	89	12 (2.6)
0.9	0.7	162	90	32 (5.1)	88	28 (5.4)	89	27 (4.6)
0.9	0.8	532	90	80 (8)	89	75 (9)	89	73 (8)
0.7	0.3	62	90	31 (4.0)	88	28 (4.3)	89	27 (4.1)
0.7	0.5	248	90	99 (7.8)	89	94 (8.2)	89	93 (8.0)
0.5	0.4	1036	90	570 (16)	89	565 (16)	89	565 (16)
0.3	0.1	158	90	126 (5.1)	89	125 (5.4)	90	124 (5.3)
0.2	0.1	532	90	452 (8)	90	451 (8)	90	451 (8)

RPW: randomized play-the-winner; DL: drop-the-loser
10 000 replications ($\alpha = 0.05$ two-sided).
The sample size was selected that yielded simulated power of approximately 90 percent under complete randomization.

of initial balls could lead to success runs that result in few of one treatment being chosen. At the other extreme, having many balls dilutes the adaptive effect. We have chosen five as a middle ground.

The drop-the-loser rule is better than the randomized play-the-winner rule in every case, having slightly larger power and fewer expected treatment failures. We see that the drop-the-loser rule preserves power quite adequately over complete randomization, and in every case results in fewer expected failures, ranging from approximately one to six fewer expected failures. While these reductions may not be dramatic, such reductions are desirable in clinical trials where treatment failures are particularly undesirable. It is clear from these results that there is little reason to use the randomized play-the-winner rule when the drop-the-loser rule is available.

Response-adaptive randomization procedures targeting optimal allocation

The question now arises, although the drop-the-loser rule is fairly successful in preserving power and reducing treatment failures, can a procedure be found that targets optimal allocation that would be more successful? The simplest form of response-adaptive randomization would be, after each patient has responded to treatment, or after a group of patients have responded, to compute the current success proportions $\hat{P}_A$ and $\hat{P}_B$ and assign the next patient to treatment A with probability (for optimal allocation)

$$P(A) = \frac{\sqrt{\hat{P}_A}}{\sqrt{\hat{P}_A} + \sqrt{\hat{P}_B}} \tag{4}$$

This procedure has been studied extensively by Melfi and Page [11] and has been called the *sequential maximum likelihood procedure* [9] and the procedure targets (or converges to) optimal allocation for moderate to large n.

One can generalize the sequential maximum likelihood procedure by the introduction of a function that measures the distance between the optimal and actual allocations at each sequential stage. This idea was originally called the doubly-adaptive biased coin design by Eisele [12]. However, the choice of an appropriate distance function has not been clear. Hu and Zhang [13] developed an appropriate function, which we now describe.

Doubly-adaptive biased coin design

This procedure can target any allocation desired, including Neyman and optimal, and is a generalization of the sequential maximum likelihood procedure. Suppose we have assigned j patients to treatment and are about to assign patient $j+1$. Let ρ be the desired allocation and let $\hat{\rho}$ be the desired allocation replaced by the current estimates $\hat{P}_A$ and $\hat{P}_B$. Let N_A/j and N_B/j be the current proportions of patients that have already been allocated to A and B, respectively. Then the probability of assigning patient $j+1$ to treatment A is given by

$$P(A) = \frac{\hat{\rho}(j\hat{\rho}/N_A)^\gamma}{\hat{\rho}(j\hat{\rho}/N_A)^\gamma + (1-\hat{\rho})((j(1-\hat{\rho})/N_B))^\gamma} \tag{5}$$

where γ is a nonnegative integer. (If $N_A/j = 0$ or 1, then $P(A) = 1$ or 0, respectively.)

To see how this procedure works in practice, we look at a simple example with $\gamma = 2$. Suppose we have already assigned 9 patients, 5 to A and 4 to B. We have observed a success rate of $\hat{P}_A = 3/5$ on A

and $\hat{P}_B = 1/4$ on B. If we are interested in optimal allocation, then we can compute

$$\hat{\rho} = \frac{\sqrt{3/5}}{\sqrt{3/5} + \sqrt{1/4}} = 0.6077.$$

Then the probability of assigning the 10th patient to treatment A is computed as

$$P(A) = \frac{0.6077(9 \times 0.6077/5)^2}{0.6077(9 \times 0.6077/5)^2 + 0.3923(9 \times 0.3923/4)^2} = 0.704. \quad (6)$$

The doubly-adaptive biased coin design satisfies our first guiding principle: it can directly target optimal allocation. For the second guiding principle, the parameter γ determines the variability of the allocation proportions arising from the randomized procedure. At the most extreme, when $\gamma = 0$, we have the sequential maximum likelihood procedure, which has the highest variability for the doubly-adaptive biased coin design. As γ tends to ∞, we have a procedure that assigns the patient to A with probability one if $N_A/j < \hat{\rho}$, and to B with probability 1 if $N_A/j > \hat{\rho}$. This procedure is entirely deterministic (except when $N_A/j = \hat{\rho}$). It turns out that the deterministic procedure has the smallest variability that can be attained by any procedure targeting optimal allocation. However, one loses the benefit of randomization, the third guiding principle. As γ becomes smaller, we have more randomization, but also more variability.

The Z-test is also asymptotically standard normal following a doubly-adaptive biased coin design using Hu and Zhang's function [13]. Simulation results have previously been reported [8] that show how well the allocation proportions approximate optimal allocation for moderate samples. The two rules based on urn models (drop-the-loser and randomized play-the-winner) are very slow to converge to urn allocation for large P_A and P_B. The doubly-adaptive biased coin design with $\gamma = 2$ tends to have very good convergence for moderate sample sizes. What have not been previously reported are simulation results for power and expected treatment failures following such a design.

We can see in Table 3 that simulation results show that $\gamma = 2$ is a good tradeoff that yields almost the same results as $\gamma = \infty$, but is slightly better than the sequential maximum likelihood procedure ($\gamma = 0$). We also simulated $\gamma = 5$ (data not shown) and it was nearly identical to $\gamma = 2$. In every case, the doubly-adaptive biased coin design with $\gamma = 2$ has the same or slightly better power than complete randomization, with a smaller expected number of treatment failures. The doubly-adaptive biased coin design is slightly more powerful than the drop-the-loser rule, but the drop-the-loser rule reduces the expected treatment failures slightly more when the success probabilities are relatively high. The doubly-adaptive biased coin design has fewer expected treatment failures when success probabilities are low.

Additional simulation results

We have examined only the expected number of treatment failures and its standard deviation. It also might be desired to compare the maximum number of failures for the different procedures we have explored. The simulated maximum numbers of failures are given in Table 4. The results are similar to the reductions seen in expected treatment failures.

Finally, an alternative way of comparing power and expected treatment failures would be to find the sample size required for a response-adaptive randomization procedure to achieve identical power to that of complete randomization, and then to compare the expected treatment failures at that sample size. While we did not do this for every case, since power was so similar to that of complete randomization (with 1 or 2% deviants in most

Table 3 Simulated power and expected treatment failures (standard deviation) for complete randomization and the doubly-adaptive biased coin design with $\gamma = 0$, 2, ∞, 10000 replications ($\alpha = 0.05$, two-sided)

			Complete		$\gamma = 0$		$\gamma = 2$		$\gamma = \infty$	
P_A	P_B	n	Power	Failures	Power	Failures	Power	Failures	Power	Failures
0.9	0.3	24	90	10 (2.4)	89	8 (2.1)	91	8 (1.7)	92	8 (1.5)
0.9	0.5	50	90	15 (3.2)	90	14 (2.9)	91	13 (2.6)	91	14 (2.5)
0.9	0.7	162	90	32 (5.1)	90	31 (4.9)	90	31 (4.8)	91	31 (4.8)
0.9	0.8	532	90	80 (8)	90	79 (8)	91	79 (8)	91	79 (8)
0.7	0.3	62	90	31 (4.0)	90	29 (3.9)	90	28 (3.5)	90	28 (3.4)
0.7	0.5	248	90	99 (7.8)	90	97 (7.6)	90	97 (7.5)	90	97 (7.4)
0.5	0.4	1036	90	570 (16)	90	567 (16)	90	567 (16)	90	567 (16)
0.3	0.1	158	90	126 (5.1)	89	122 (5.5)	90	122 (5.4)	90	122 (5.3)
0.2	0.1	532	90	452 (8)	90	448 (9)	90	448 (9)	90	448 (8)

The sample size was selected that yielded simulated power of approximately 90 percent under complete randomization.

Table 4 Simulated maximum number of treatment failures for complete randomization, the randomized play-the-winner (RPW) rule, the drop-the-loser (DL) rule, and the doubly-adaptive biased coin design (DBCD) with $\gamma = 2$, 10000 replications ($\alpha = 0.05$, two-sided)

P_A	P_B	n	Complete	RPW	DL	DBCD
0.9	0.3	24	18	17	15	15
0.9	0.5	50	28	26	25	24
0.9	0.7	162	51	50	47	48
0.9	0.8	532	115	111	107	112
0.7	0.3	62	45	42	43	42
0.7	0.5	248	129	126	126	126
0.5	0.4	1036	633	632	623	631
0.3	0.1	158	145	146	143	143
0.2	0.1	532	482	482	479	480

The sample size was selected that yielded simulated power of approximately 90 percent under complete randomization.

cases), we give an illustration of this method of comparison for $P_A = 0.9$ and $P_B = 0.8$. For this set of parameters, we have 90 percent power under complete randomization at $n = 532$, and 80 treatment failures are expected (standard deviation = 8). Under the randomized play-the-winner rule, we would require $n = 554$ for identical power, resulting in 78 expected treatment failures (standard deviation = 9). For the drop-the-loser rule, we require $n = 540$ with 74 expected treatment failures (standard deviation = 8). For the doubly-adaptive biased coin design with $\gamma = 2$, we require only $n = 528$, but with 78 expected treatment failures (standard deviation = 8). So for this set of parameter values, the drop-the-loser rule requires 12 more patients than the doubly-adaptive biased coin design, but results in four fewer treatment failures.

Conclusions

We have discussed response-adaptive randomization procedures that are useful for very simple clinical trials comparing two treatments with respect to binomial success probabilities. Similar research can be extended to trials with continuous responses or survival outcomes, and this is ongoing work by the authors of this paper. Only a minority of clinical trials fit into the context of this paper, as many clinical trials have long-term outcomes and limited recruitment in the beginning of the trial, thus not allowing for adapting on patient responses. However, the procedures discussed here are relevant when there are delays in response, as the procedure can be updated when information becomes available, or as groups of patients respond. However, delayed response, particularly if responses occur too far out in the randomization scheme, can effectively dilute the beneficial impact of the procedure.

So for the minority of clinical trials for which these procedures can be reasonably implemented, our simulations demonstrate that the doubly-adaptive biased coin design is, for the parameter values we examined, always as powerful or slightly more powerful than complete randomization, with fewer expected treatment failures. The drop-the-loser rule has similar results, and in some cases better results in terms of expected failures, with losses of no more than one percent in power. It is clear from our simulation results that there is no situation we examined for which the randomized play-the-winner rule is preferable to the drop-the-loser rule.

Let us more closely examine the drop-the-loser rule and the doubly-adaptive biased coin design to determine which one should use in practice. It appears that results are quite similar when the study is powered to detect large differences. For small differences, if the success probabilities are large, the drop-the-loser rule reduces the expected failure rate more than the doubly-adaptive biased coin design. When the treatments are not very successful, say less than 0.5, the doubly-adaptive biased coin design has a lower expected failure rate. When a study is planned, information on the success rates that are desired to be detected should therefore be used to select the appropriate procedure. Sample size requirements can be computed by simulation (the first author has simple SAS programs for each procedure available upon request), and in the case of the doubly-adaptive biased coin design, the sample size required may be slightly smaller than that required for complete randomization.

If the clinical trial fits into the scenario presented in this paper, is there any reason why one would wish to use equal allocation when the drop-the-loser rule and doubly-adaptive biased coin design are available? Essentially the usual test statistic can be used, sample size computations can be done at the beginning of the trial, and the only change is that randomization will be performed slightly differently. The benefits will be that the study will be sufficiently powered while making modest reductions in expected treatment failures. So we believe the answer is generally 'no', but there are certain issues that need to be addressed. For example, the informed consent document would have to incorporate information on the randomization scheme, and some suggestions on appropriate ways to do this are found in Chapter 12 of [9]. Other details on bias, masking, and other logistical considerations are discussed in that chapter.

Modest reductions in expected treatment failures will obviously be more important in clinical trials with serious outcomes. However, one important

point to take into consideration is that in studies powered for a difference that is smaller than the difference actually observed, power will increase, and a response-adaptive randomization procedure will be even more favorable in terms of the expected success rate than complete randomization.

Acknowledgements

This research was supported by grant DMS-0204232 from the National Science Foundation.

References

1. **Rosenberger WF, Stallard N, Ivanova A, Harper C, Ricks M.** Optimal adaptive designs for binary response trials. *Biometrics* 2001; **57**: 909–13.
2. **Friedman LM, Furberg CD, DeMets DL.** *Fundamentals of clinical trials*. Boston: Wright PSG, 1981.
3. **Berry DA, Fristedt B.** *Bandit problems: sequential allocation of experiments*. London: Chapman and Hall, 1985.
4. **Berry DA.** One- and two-armed bandit problems. In Kotz S, Johnson N eds, *Encyclopedia of Statistical Sciences*, Vol 6. New York: John Wiley, 1985, 418–22.
5. **Wei LJ, Durham S.** The randomized play-the-winner rule in medical trials. *JASA* 1978; **73**: 840–43.
6. **Ivanova AV.** A play-the-winner type urn model with reduced variability. *Metrika* 2003; **58**: 1–14.
7. **Rosenberger WF, Flournoy N, Durham SD.** Asymptotic normality of maximum likelihood estimators from multiparameter response-driven designs. *Journal of Statistical Planning and Inference* 1997; **60**: 69–76.
8. **Hu F, Rosenberger WF.** Optimality, variability, power: evaluating response-adaptive randomization procedures for treatment comparisons. *JASA* 2003; **98**: 671–78.
9. **Rosenberger WF, Lachin JM.** Randomization in clinical trials. New York: Wiley, 2002.
10. **Royall RM.** Ethics and statistics in randomized clinical trials. *Stat Sci* 1991; **6**: 52–62 (with discussion).
11. **Melfi V, Page C.** Estimation after adaptive allocation. *Journal of Statistical Planning and Inference* 2000; **87**: 353–63.
12. **Eisele JR.** The doubly adaptive biased coin design for sequential clinical trials. *J. Statist. Plann. Inf.* 1994; **38**: 249–61.
13. **Hu F, Zhang L-X.** Asymptotic properties of doubly adaptive biased coin designs for multi-treatment clinical trials. *Ann Stat* 2004; **30**: in press.