Response Adaptive Designs for Balancing Complex Objectives

Janis Hardwick and Quentin F. Stout

University of Michigan
Ann Arbor, Michigan 48109

Abstract

Response adaptive designs for problems involving multiple conflicting goals are considered. The often referenced clinical trial dilemma of trying to determine the better of two independent Bernoulli populations while simultaneously increasing successful results during the trial is used. In evaluating the continuum of performance tradeoffs between the two objectives it is shown that only minor decreases in the efficiencies of the objectives are needed to obtain nearly optimal performance on both. A Bayesian model is used and the analysis is carried out by combining both objectives into a single objective function and using dynamic programming to optimize the result. By varying the relative weights of the objectives one can see the entire range of optimal tradeoffs possible. A number of ad hoc approaches are also evaluated and it is shown that a modified 2-armed bandit strategy exhibits the best behavior when it comes to balancing these objectives. Pointwise examinations of the operating characteristics of all designs are also considered.

Keywords: sequential design, bandit problem, randomized play-the-winner, clinical trial, optimal tradeoff, Gittins index, multiple objectives

1 Introduction: Adaptive Designs

A design is a procedure or algorithm that specifies how to allocate resources during a study. It is considered to be good or even optimal if it allows for sufficiently precise and accurate data analysis with the least output of costly resources. Most experimental designs use fixed sampling procedures in which the sample sizes and order of allocations to different populations are known in advance. Such sampling rules are easy to apply and intuitive. Occasionally, one can analytically determine the optimal fixed sampling procedure. More often, however, the most efficient procedure for allocating among populations is a function of unknown parameters.

Adaptive designs allow investigators to adjust resource expenditures while the experiment is being carried out. In this way, accruing data can be used to estimate parameters and guide future allocations. In some experiments, adaptation depends merely on knowledge of the allocations themselves. This may occur, for example, when attempting to balance treatment

groups across covariates. Usually, however, adaptation also incorporates knowledge of the responses observed to date. Such designs are referred to as *response adaptive* designs.

For most response adaptive problems it has been infeasible to derive allocation procedures that generate experimental designs that are optimal in the desired sense. Often the evaluation of such designs has been based on asymptotic approximations [2, 24, 31]; although, simulation techniques have also been popular evaluation methods [7, 16]. A more recent approach has been to utilize computer algorithms to determine the sampling designs as well as to evaluate them [19, 22, 23]. The combination of improved algorithm implementation and continued increases in computational power make the generation of adaptive designs possible. In this paper, we use *exact* computational methods for both the generation and the evaluation of all designs considered, along with their operating characteristics. In other words, the solutions presented here are not approximations for any of the techniques discussed.

One difficulty with the pursuit of optimal designs is that investigators typically have several experimental objectives. Often then, good performance with respect to one objective works against the performance of others. In such situations, it is desirable to understand how such competing objectives make these performance tradeoffs. Examples of potentially important objectives are, among others, overall financial cost, time to decisions, quality of decisions and error rates, subject well-being both during and after the experiment, robustness to departures from assumptions, as well as distributions of such characteristics. The goals of this paper are twofold. The first is to determine the exactly optimal tradeoff curve for two competing objectives. The second is to illustrate how well various popular allocation procedures perform in this context. Both goals represent new work. An exactly optimal tradeoff curve of this nature has never been calculated. Further, the exact analysis of the performance of the various suboptimal designs is almost entirely new.

As an example to illustrate the techniques described, we consider a popular clinical trials model in which two important objectives compete for experimental resources. The first objective is to determine the better of two Bernoulli populations at the termination of an experiment of size N. The second is to design the experiment such that patient treatment within the study is as successful as possible. This problem represents the natural tension between efficiency and patient care since reaching a good decision will require extensive sampling on each treatment arm, while good patient care argues for exploiting the arm that appears to be best. Naturally, there are many other potentially conflicting objectives for the k Bernoulli population scenario. Fortunately, the techniques discussed here can usually be applied in these other cases as well.

The paper is organized as follows. In Section 2, optimization models and the use of dynamic programming to solve them are introduced. In Section 3, some common suboptimal designs are described along with a design based on approximating the Gittins index for infinite geometrically discounted bandit optimization. In Section 4, the specific objectives to be compared are discussed along with two efficiency measures. The definition and generation of the optimal tradeoffs between these measures are also presented. In Section 5 design performance is examined as well as certain pointwise operating characteristics. Section 6 contains some final remarks.

2 Models for Optimization

The term "optimal" can be somewhat confusing. First of all, fully optimal designs for models with unknown parameters must be omniscient and are thus unobtainable in practice. Next, except in trivial cases, no design is uniformly optimal for all parameter configurations. As a result, restrictions are placed on a class of designs and a well-defined objective is the focus of design development. Optimal designs are those that are the best obtainable only in reference to this restricted class and specified objective. As examples, frequentists often look for designs that are mini-max or have minimum mean squared error with respect to the study objective. Bayesians seek the best design given that the results are integrated with respect to a prior distribution on the parameter space. In a research field often referred to as alphabetic optimization, various measures of imprecision based on the information matrix of the parameters are used as optimality criteria [1, 10]. These measures, which include A-, C-, D-, V- among others, are heavily model dependent but cover a variety of important optimality standards.

Here we define a design to be exactly optimal if, in the Bayesian sense and for a specified prior, there is no better design to be found for the given objective unless the parameters are known in advance. While the focus of this paper is competing objectives, even the simpler goal of exactly optimizing adaptive designs for a single objective has repeatedly been viewed as intractable.

As an example, in 1956, in [8] the authors argue that if, for a specific problem, the optimal sequential procedure were "practically obtainable, the interest in any other design criteria which have some justification although not optimal is reduced to pure curiosity." While we disagree with this perspective, for our purposes what is interesting about this comment is that it is immediately followed with the statement that actually obtaining optimal procedures "is not practicable". Then, as an illustration of the "intrinsically complicated structure" of optimal procedures, the authors detail the first step of the optimal solution to a simple sequential design problem involving only three Bernoulli observations. During this same year, it was pointed out in [5] that problems of this nature could, in principle, be solved via *dynamic programming*. While the previous arguments were valid at the time of those writings, such solutions still tend to be viewed as infeasible. Thirty five years later, in [32], when addressing a variation of the problem in [8], the author reiterates the view that "In theory the optimal strategies can always be found by dynamic programming but the computation required is prohibitive". Similarly in [1], the authors note that "The construction of an optimum sequential scheme depends on the relative costs of experimentation and of time lost in obtaining an accurate answer. In all but the simplest cases the calculations are rarely performed".

Despite the manifest difficulties associated with determining and possibly implementing fully optimal designs, when they *can* be obtained, such designs provide a gold standard that illuminates the entire design development process. In the next section, we describe the exact optimization procedure known as dynamic programming.

2.1 Dynamic Programming

Dynamic programming is an example of an algorithmic technique that has flourished with the advent of refined implementations and superior computers. Throughout the decision sciences and computer science, dynamic programming has been proven to be a powerful optimization technique [5]. Here we describe it for the fully sequential or adaptive Bayesian case, with other Bayesian problems having similar developments.

Suppose an objective function $\mathcal{O}$ is defined on the terminal states of the experiment, and the goal is to maximize the expected value of $\mathcal{O}$. We assume that the sampling options available, and responses obtained, are discrete. During the experiment, suppose we are at state σ and can sample from populations $P_1, \ldots, P_k$. For population P_i, suppose that at state σ there are $r(i)$ possible outcomes, $o^i_1, \ldots, o^i_{r(i)}$, occurring with probability $\pi^i_1(\sigma), \ldots, \pi^i_{r(i)}(\sigma)$. Let $\sigma + o^i_j$ denote the state where o^i_j has been observed by sampling P_i. Let $\mathcal{E}_{opt}(\sigma)$ denote the expected value of the objective function attained by starting at state σ and sampling optimally, and let $\mathcal{E}^i_{opt}(\sigma)$ denote the expected value of the objective function attained by starting at state σ, observing P_i, and then proceeding optimally. These expectations are taken with respect to the Bayesian model in which we have a joint prior distribution, Φ, on the $\pi^i_j(\emptyset), i = 1, \ldots, k,\ j = 1, \ldots, r(i)$, where $\emptyset$ is the initial state of the experiment. Then $\pi^i_j(\sigma) = E^{\Phi}(\pi^i_j(\emptyset) \mid \sigma)$ are the posterior means given state σ.

The critical recursive relationship, sometimes referred to as the principle of optimality, is that

$$\mathcal{E}^i_{opt}(\sigma) = \sum_{j=1}^{r(i)} \pi^i_j(\sigma) \cdot \left[\mathcal{E}_{opt}(\sigma + o^i_j)\right]. \tag{1}$$

Since the only actions available are to stop with value $\mathcal{O}(\sigma)$ or to sample one of the populations, we have

$$\mathcal{E}_{opt}(\sigma) = \max\left\{\mathcal{O}(\sigma),\, \max\{\mathcal{E}^i_{opt}(\sigma) : i = 1, \ldots, k\}\right\} \tag{2}$$

The dynamic programming algorithm starts at the terminal states, and then for each of their predecessor states determines the action that will optimize the expected value.

Unfortunately, the elementary nature of equation 2 is misleading. Even using the best approaches known, the simplest k-population problem with d outcomes is daunting, as its computational requirements grow as $N^{dk}/(dk{-}1)!$, where N is the sample size. As other constraints and costs are added [19, 23], such problems quickly outgrow the capabilities of PCs. Thus we are sometimes forced to utilize parallel computing [30]. Still, ongoing development of new algorithms and more precise implementations allow us to push computational limits so that increasingly complex problems in adaptive sampling can be solved.

Note, however, that not all important problems have objective functions satisfying recursive equations. For example, many mini-max objectives are not defined in terms of expectations with respect to a distribution on the populations, but rather as a maximum or minimum over the populations.

2.2 Multi-arm Bandit

A popular class of problems optimized by dynamic programming are finite multi-arm bandit problems. The name comes from the following illustration. Suppose that you have a slot

machine with k arms and N tokens. If you deposit a token to pull Arm i, you receive \$1 with an unknown probability p_i and \$0 with probability $(1 - p_i)$, $i = 1, \ldots, k$. The arms are independent and your goal is to sample from them sequentially in such a way as to maximize your winnings after the N pulls. Bandits are commonly used as models for learning, scheduling, searching, screening and economics [3, 6, 14], where the desire for immediate yield counterpoises that for future rewards, the magnitude of which may be revealed only after additional information has been gathered. In particular however, from their earliest inception multi-armed bandits have been used to model "ethical" clinical trials [26]. In this context, the generic bandit goal of optimizing payoff, which may be translated as subject well-being, fits in well with typical goals of patient care. However, in a clinical trial investigators want to treat each subject in the experiment as well as possible, but must balance this against the longer term aim of the trial, which is to collect enough evidence to make informed decisions about how best to treat future patients. Thus, a bandit scenario makes a reasonable model for our example with the competing goals of good decision making and good patient treatment during the trial. Note however, that the objective that the bandit solution explicitly optimizes is the latter goal. In the present case, this is maximizing successful outcomes during the experiment. Since dynamic programming is used to optimize many objectives, we refer to the dynamic programming solution that optimizes this particular bandit design as "OS" which stands for "optimizing successes".

3 Classes of Suboptimal Solutions

In some situations, optimal procedures may not be used because they are inaccessible, complex, or difficult to employ and explain. Still, as noted, they provide a basis of comparison to establish the efficiency of suboptimal designs. When the relative efficiency of a sampling procedure is high compared with the optimal one, then investigators may be justified in implementing a simpler and, typically, more intuitive suboptimal option. There are several important classes of suboptimal solutions and in some cases they may be the only procedures available. Thus, it is important to repeatedly evaluate suboptimal techniques for the various problems for which optimal designs *do* exist. A set of such comparisons generates a prospective guide for investigators in selecting the suboptimal procedure most well suited for their particular problem.

What follows are descriptions of a variety of potentially useful classes of experimental designs.

Myopic Allocation: At each stage of an experiment, *myopic* or *greedy* rules use the choice that would be optimal if there was only one observation left. Note that this is the approach you would prefer if you were the next patient in the clinical trial. Such rules are intuitive and often analytically tractable. For this reason, significant research has focused on problems in which myopic strategies prove to be optimal or asymptotically optimal [11, 12, 36]. For many other problems however, it is unlikely that myopic rules will perform very well because such rules cannot adequately incorporate certain complexities. In particular, if there is significant uncertainty about the values of the success rates for the treatment arms, then it may be that only a few successes on one arm cause a myopic rule to repeatedly choose that arm,

without having gathered much information about the others. Such premature convergence may result in permanently picking an arm that is significantly inferior.

Here we analyze a Bayesian myopic strategy, where the priors are the same as the design priors for the optimal allocation. We use a tie-breaking rule whereby if two populations have the same expected value then we select the one for which there have been fewer observations. If they have the same number of observations then we randomize. We refer to the myopic design as "MY".

m-Stage Look-ahead Rules (m-sla): A generalization of the myopic strategy is the m-stage look-ahead rule that incorporates information further into the future but not to the end of the horizon. Solutions for m-sla rules are obtained by repeatedly locating the dynamic programming solution for the problem ending m-stages ahead, $1 \leq m \leq N$. Naturally, as m approaches N, this strategy converges to the optimal solution for an N-horizon problem, and when $m = 1$, the rule is myopic. Quite productively, m-sla rules are utilized as approximately optimal solutions to stopping rule problems [11].

Hyperopic Allocation: Many experiments use sampling proportions that are fixed in advance. While the most popular such rule is equal allocation (EA), optimal fixed sampling designs are derived more generally through the optimization of an objective function such as a risk function. In solving such problems, one reduces the curse of dimensionality associated with dynamic programming; and, while the calculations can be quite difficult, often requiring multi-dimensional numerical integration, they are typically manageable.

A sequential version of the "best fixed" rule is to re-solve the fixed sample size problem at each stage of the experiment using the information that has been taken in so far. We refer to such strategies as *hyperopic* because at each stage they look to the end of the experiment and re-determine the best fixed sample sizes. Various heuristics based on the new best fixed rule may be used for the subsequent allocation. Hyperopic allocation rules have performed surprisingly well in a number of applications [19, 20, 23], although with each application the heuristic must be adapted to fit the new problem objective.

Urn Models: Numerous response adaptive designs proposed in the literature have been based on *urn* models [9, 13, 24]. The general idea is to imagine an urn filled with balls of different types that represent the different populations. At each stage, a ball is selected at random and the corresponding population is sampled. Depending on the outcome of an observation, a non-negative number of balls (not necessarily discrete) of each type are added to the urn. Sampling may be with or without replacement. Generally speaking, performance of urn models is based on first order convergence properties of the stochastic process generated by continued sampling from the urn. The model discussed most often is the *randomized play the winner* rule (RPW) [2, 7, 27, 34, 35]. In the RPW, one starts with an urn containing α_i balls of type i for $i = 1, \ldots, k$. Sampling is with replacement, and if the last response is a success on Arm i, then β_s balls of type i are placed in the urn. If a failure occurs, then β_f balls from the other populations are added. Most often, $k = 2$, $\alpha_1 = \alpha_2 = 1$ and $\beta_s = \beta_f = 1$. These are the values used in this paper.

A Modified Bandit: Recall that the experimental model described in Section 2.2 is optimized via dynamic programming for the objective of obtaining the most successful outcomes in

an experiment. That particular bandit model is known as the *finite horizon uniform 2-arm Bernoulli bandit*. In the more general bandit scenario, the populations have unknown reward structures and allocation decisions are made with the goal of optimizing a "discounted" sum of all observations which may even be infinite.

In trying to develop a design that performs well on both of our objectives, it is useful to note that the finite multi-arm bandit solution becomes myopic near the end of a trial. The procedure tries to maximize successes while focusing less on the goal of making a good decision that will apply to post-trial patients. Thus, seeking designs that place value on obtaining successful outcomes for patients both during and after the trial can be advantageous in balancing our two goals.

There is, for example, one bandit model that is extremely close to the uniform finite horizon model but which also looks further into the future. This is the *geometric* bandit in which the i^{th} outcome is discounted by β^{i-1}, $i = 1, \ldots, \infty$, for $0 < \beta < 1$. Values of β close to one maintain an emphasis on information gathering whereas smaller values force the bandit to behave more myopically. Thus the selection of the parameter β is an important part of the design specification.

Since the sample size is infinite, dynamic programming is not practicable. However, with geometric discounting and independent arms, optimal solutions can be defined in terms of the Gittins index rule [15]. For every state in the experiment, for every arm an index exists that depends only on the arm. The optimal strategy is to sample from the arm associated with the highest index.

Despite the elegant theoretical solution imparted by the Gittins index theory, the indices are nevertheless complicated and extremely difficult to calculate, even in very simple models. In most cases they cannot be calculated exactly. Still, the idea of an index rule having traits similar to the Gittins index is very appealing. For the case in which the reward functions are Bernoulli with parameters that follow beta distributions, an index rule based on a lower bound that closely approximates the Gittins index is proposed in [18]. It is known as the "modified" bandit rule or "MB". If, for a given arm, the parameters for the beta prior are A and B, then a lower bound for the Gittins index for this arm is given by $\Lambda^* = \sup\{\Lambda_s : s = 1, 2, \ldots\}$, where

$$\Lambda_s = \frac{\frac{\Gamma(A+1)}{\Gamma(A+B+1)} - B\Sigma_1^s \beta^i \frac{\Gamma(A+i)}{\Gamma(A+B+i+1)}}{\frac{\Gamma(A)}{\Gamma(A+B)} - B\Sigma_1^s \beta^i \frac{\Gamma(A+i-1)}{\Gamma(A+B+i)}} \ .$$

Generally speaking, it is not difficult to locate Λ^* since $\{\Lambda_s : s = 1, 2, \ldots\}$ is unimodal. Let $s^* = \min\{s : \Lambda_{s+1} \leq \Lambda_s\}$, and simply take $\Lambda^* = \Lambda_{s^*}$.

4 Tradeoffs and Evaluation

As noted in Section 1, combinations of desirable study features compete and must be balanced against one another in the study design. While balance is often examined and occasionally designed for, attempts to generate optimal tradeoff curves have rarely if ever been addressed. Recall that the focus here is to improve patient outcomes during a clinical trial

while also gathering information to make a good terminal decision regarding the treatment of future patients, see [25].

In order to assess tradeoffs between these objectives, each objective must have an associated performance measure. Suppose there are k populations with success rates p_i for population i, $1 \leq i \leq k$, and let $\eta(p_1, \ldots, p_k)$ be the joint prior on the p_i's. Also, let $p^* = \max_i p_i$. Then, given a design δ, let $\mathrm{E}^\delta(n_i)$ be the expected number of observations on population i when there are a total of $\mathrm{E}^\delta(N)$ observations. Here N is expressed as a random variable because δ may incorporate a stopping rule. In our evaluations, we use the following two efficiency measures: $\mathcal{S}$ for patient well-being, and $\mathcal{D}$ for good terminal decisions.

We define *sampling efficiency* as:

$$\mathcal{S}(\eta, \delta) = \mathbf{E}^\eta \left[\frac{\sum_{i=1}^k p_i \, \mathrm{E}^\delta(n_i)}{p^* \, \mathrm{E}^\delta(N)} \right].$$

This performance measure is closely related to *expected successes lost:* $p^*\mathrm{E}^\delta(N) - \sum_{i=1}^k p_i \mathrm{E}^\delta(n_i)$, see [4], which is also known as *regret*. Note that expected successes lost is an absolute measure of success rates, as opposed to $\mathcal{S}$ which measures them relative to p^*.

Next, if we have a decision rule at the end that selects the treatment to be preferred for future patients, then given δ, let $\xi_\delta(i)$ be the probability that population i is selected, and define *decision efficiency* as:

$$\mathcal{D}(\eta, \delta) = \mathbf{E}^\eta \left[\left\{ \Sigma_{i=1}^k \xi_\delta(i) p_i \right\} / p^* \right].$$

For simplicity, we tend to drop the (η, δ) notation from $\mathcal{S}(\eta, \delta)$ and $\mathcal{D}(\eta, \delta)$. The decision efficiency measure is also related to a common performance measure — the *probability of correct selection*. When treatment i is selected, then $\mathcal{D}$ penalizes incorrect decisions less when $(p^* - p_i)$ is small. This is because future patients given such a treatment will not be harmed as much as when $(p^* - p_i)$ is large. In contrast, the probability of correct selection measure treats these two cases identically. Note that $p^*\mathcal{S}$ is the expected success rate for patients in the experiment and $p^*\mathcal{D}$ is the same rate for subsequent patients if they are given the treatment selected as best.

As mentioned, no procedure is optimal with respect to all $(p_1, \ldots, p_k)$ sequences. To examine $\mathcal{S}$ versus $\mathcal{D}$ for the different procedures, we can utilize either Bayesian or frequentist versions of $\mathcal{S}$ and $\mathcal{D}$. The frequentist versions simply exclude the information from the prior distribution in the final parameter estimates. We also examine $\mathcal{S}$ and $\mathcal{D}$ broken down pointwise to understand important operating characteristics of the designs.

4.1 Optimal Tradeoffs

In the next section, we present the optimal tradeoff curve for $\mathcal{S}$ versus $\mathcal{D}$ for specific priors and sample sizes. That is, among all procedures that achieve a given value of $\mathcal{S}$, we determine the one that maximizes $\mathcal{D}$. We also determine the largest values of $\mathcal{S}$ and $\mathcal{D}$ that can be achieved.

This analysis is carried out by maximizing the objective function $\alpha\mathcal{S} + (1-\alpha)\mathcal{D}$, for $0 \leq \alpha \leq 1$. Dynamic programming is used to determine the design $\mathcal{A}_\alpha$ that optimizes this function. Then $\mathcal{A}_\alpha$ is evaluated to determine its $\mathcal{S}(\eta, \mathcal{A}_\alpha)$ and $\mathcal{D}(\eta, \mathcal{A}_\alpha)$ values. For each $\alpha \in [0, 1]$, this provides a point on the optimal tradeoff curve.

5 Analysis of Procedures

For the present problem, $\mathcal{S}$ and $\mathcal{D}$ represent performance measures of two features critical for a good design. For high $\mathcal{D}$ it is essential to obtain efficient estimates of (p_1, p_2). This requires the procedure to continue sampling from both populations even if one initially looks better, and to sample more from the one having highest variance. Alternatively, for high $\mathcal{S}$, there must be an emphasis on sampling more observations from the better population. This exploitation of immediate gain works against the exploration needed for high $\mathcal{D}$.

In what follows we present the optimal tradeoff curve for $\mathcal{S}$ versus $\mathcal{D}$. We also provide evaluations of various ad hoc rules and analyze their behaviors. These designs are

- Equal Allocation (EA) to represent fixed allocation;
- Randomized Play-the-Winner (RPW) to represent urn allocation;
- Myopic Allocation (MY) to represent rules that adapt but are short sighted;
- Uniform 2-arm Bandit (OS) achieved by dynamic programming to optimize successes;
- Modified Bandit (MB) using Gittins index approximations with β values chosen to perform well on both $\mathcal{S}$ and $\mathcal{D}$.

Note that for space considerations we do not analyze the m-stage look-ahead and hyperopic designs here. They were described as part of a continuum of designs with various look-ahead properties but, predictably, their performance falls within the extremes of the designs that are analyzed in what follows.

We use the natural decision rule of selecting the population corresponding to the highest frequentist estimate of the population success rate. In the case of ties, each is selected with equal probability. All calculations in this section are computed exactly and no simulation is used. The pointwise distributions are obtained using a technique known as *path induction*, see [21]. We examined sample sizes of 20, 50, 100 and 150 but for space considerations we use $N = 50$ for most of the analyses. For each procedure we began with one allocation from each population. As defined previously, this is not a requirement of any procedure except EA. However, it removes the difficulty of making a decision when there are no observations from one population.

The tradeoffs for the uniform case when $N = 50$ are illustrated in Figure 1(a). The continuous curve in the figure represents optimal tradeoffs between $\mathcal{S}$ and $\mathcal{D}$. Let $\mathcal{S}_{opt}$ and $\mathcal{D}_{opt}$ represent the extremal points corresponding to the weights $(\alpha, 1-\alpha)$ of (1,0) and (0,1) on $(\mathcal{S}, \mathcal{D})$ respectively. These two extremes are the points leftmost for (0,1) and rightmost

for (1,0) on the optimality curve. While it would be up to investigators to decide how to weigh the two efficiency criteria, one natural point of interest is the configuration where $\mathcal{S}$ and $\mathcal{D}$ have equal weight. In Figures 1(a) and 1(b), this point is signified by "=". Note that visually this point moves along the curve as a function of the ranges of the two axes. The axes are presently set to display the data as fully as possible within the graph. The $\mathcal{S}$ axis ranges from 0.725 to 0.96 whereas the $\mathcal{D}$ axis ranges only from 0.94 and 0.99. Thus, for the most part, differences among points on the $\mathcal{D}$ axis are much smaller than those for $\mathcal{S}$. However, if the decision determines which treatment is applied to future patients, then even small differences will be significant if the number of future patients is $\gg N$.

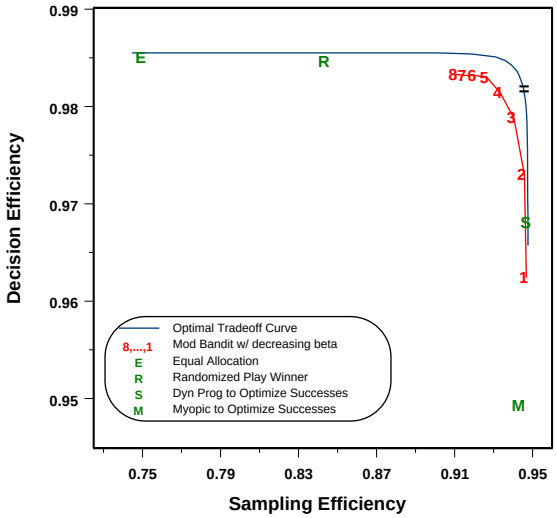


Figure 1(a) $N = 50$

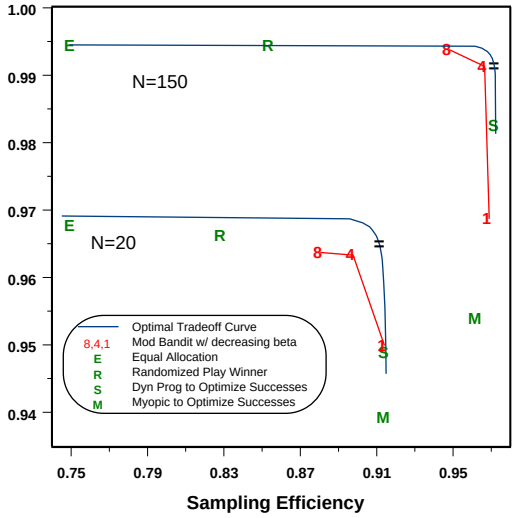


Figure 1(b) $N = 20, 150$

Once investigators have determined their desired location on the optimality curve, they can either elect to use the optimal procedure, or select a suboptimal allocation procedure with $(\mathcal{S}, \mathcal{D})$ close to this point. This process is aided by viewing $(\mathcal{S}, \mathcal{D})$ for the other methods evaluated.

Ad hoc Designs:

Myopic Because of the possibility of premature convergence, asymptotically (in N) the myopic design MY will not converge to 1 on either measure. For example, it might initially observe a success on Arm 1, a failure on Arm 2, and then never sample Arm 2 again, even though it might in fact be superior. At the end MY would decide in favor of Arm 1, and thus have limited decision efficiency as well.

Equal Allocation Note that, as seen in Figure 1(b), EA has a fixed sampling efficiency that does not improve as the sample size increases. It has high decision efficiency, but it might be surprising to note that it does not have the optimal decision efficiency and is thus ad hoc with respect to both performance measures. For example, suppose that on each arm one success and one failure have been observed and that two observations remain. If Arm 1 has a very sharp prior with a mean success rate of 1/2, while Arm

2 has a very weak prior (perhaps uniform) with the same mean, then the decision efficiency is optimized by having the last two observations on Arm 2, rather than one on each arm. This suboptimality can clearly be seen in Figure 1(b) for $N = 20$.

Randomized Play the Winner Because the standard RPW rule puts in balls of the opposite arm whenever the sampled arm has a failure, it will converge to sampling a fixed proportion of the time from each arm. This aspect makes it behave somewhat similarly to equal allocation since asymptotically it will not converge to the optimal $\mathcal{S}$, but the repeated sampling on each arm will cause convergence to $\mathcal{D} = 1$.

Bandit The success optimizing bandit, OS, lags EA and RPW in decision efficiency because it is more likely to stay on an arm that has been succeeding, rather than repeatedly sampling the other arm. While it will do such sampling, it is at a far lower rate, and thus it is slightly less likely to discover that the first few observations were misleading. Although the OS optimizes successes, it does not quite optimize $\mathcal{S}$. A counter example is as follows: suppose that one arm is known to have a success rate of 0.5, while the second has a uniform distribution. If there is only one observation left, then sampling the first arm gives an expected sampling efficiency of $\int_0^{1/2} 1\,dx + \int_{1/2}^1 0.5/x\,dx \approx 0.85$, while sampling the second gives $\int_0^{1/2} x/0.5\,dx + \int_{0.5}^1 1\,dx = 0.75$. $\mathcal{S}$ is a continuous function of the known rate, so even if the first arm has a known success rate a bit smaller than 0.5 it would still be best to observe it to increase $\mathcal{D}$. This is the case despite the fact that the second arm would have a higher mean success rate and thus would have been chosen by OS, which is myopic when there is only one observation left.

Modified Bandit The MB design was developed to reduce the myopic behavior of OS towards the end of the trial by introducing the discounted value of an infinite stream of observations, i.e., implicitly considering the patients after the trial. For fixed N, as β increases, $\mathcal{D}$ increases, along with a slight decrease in $\mathcal{S}$. This is visible in Figure 1(a) which shows $\mathcal{S}$ and $\mathcal{D}$ for a monotonic sequence of β's. These MB designs are indicated by the number of nines in the decimal β. For example, the number 4 on the plot represents $\beta = 0.9999$. In this figure β varies from 0.9 to 0.99999999 and the corresponding designs are denoted in the text as MB1 to MB8. To achieve good behavior for this problem, all of the β values selected are very close to 1, with the higher values corresponding to increasing concern about the future. On the other hand, for fixed N, as β decreases towards zero, $\mathcal{S}$ and $\mathcal{D}$ converge to the respective values attained by the myopic rule.

One striking feature of the graphs is that the optimal tradeoff curve shows that there many optimal designs that are extremely good on both measures simultaneously, giving up a little on each to gain significantly on the other. Another salient result is that several MB designs also strike such a compromise. With the proper choice of β these designs can be quite close to the desirable upper left part of the optimal tradeoff curve.

To illustrate how performance changes as a function of the sample size, Figure 1(b) provides $\mathcal{S}$ and $\mathcal{D}$ for both the optimal and suboptimal designs when $N = 20$ and $N = 150$.

Once again the MB makes very good tradeoffs. In particular, for this sample range, the value of MB4 is consistently the point on the plots that is closest to the location of "=". This means that there is a good range of sample sizes for which one need not change β to obtain a design nearly as good as the optimal one that places equal weights on the two efficiency measures.

Also of interest is how close MB8 is to having optimal $\mathcal{D}$ when $N = 150$. While in the figure, EA and RPW appear to be "on" the optimality curve, they both have $\mathcal{D}$ values slightly less than optimal. However, the four designs MB4, MB8, EA and RPW match the the optimal $\mathcal{D}$ value of 0.99 out to two decimal places. It is also the case that the three designs MB4, MB1 and OS match the optimal $\mathcal{S}$ of 0.97 out to two places.

When $N = 20$ the MB designs do not perform as well on $\mathcal{D}$ as they do when N is larger. While MB8 still has high $\mathcal{D}$, this sample size is too small to allow the exploration needed while still obtaining good immediate payoff in terms of successes. EA and RPW still perform quite well on $\mathcal{D}$. With regard to $\mathcal{S}$ however, even when $N = 20$, three designs are almost perfect. OS, MB1 and MY all match the optimal $\mathcal{S}$ of 0.91 out to two places.

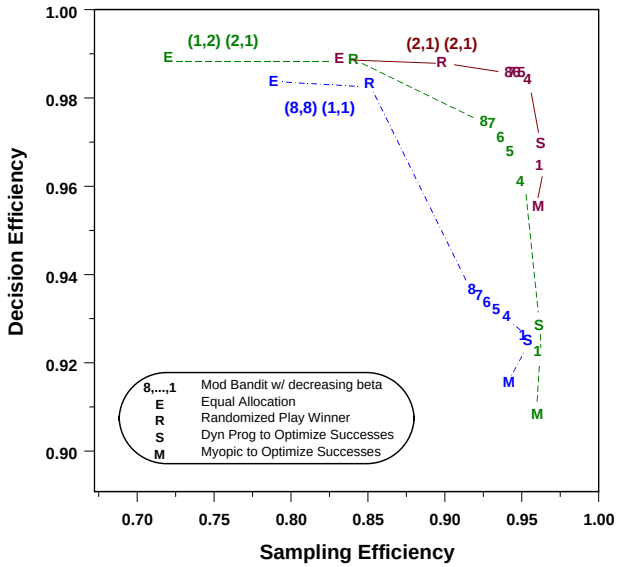


Figure 2 $N = 50$, multiple priors

Non-uniform Priors:

We also examined a range of independent beta prior distributions. In Figure 2 we include only three variations to save space. However these three sets of prior distributions exhibit roughly the same types of behavior shown by the others examined. While the optimal tradeoff curves are not included, they are similar to those in Figures 1(a) and (b). Figure 2 represents $N = 50$ and the three priors displayed are

$$p_1 \sim Be(1,2)\,,\ p_2 \sim Be(2,1); \quad p_1 \sim Be(2,1)\,,\ p_2 \sim Be(2,1);$$
$$p_1 \sim Be(8,8)\,,\ p_2 \sim Be(1,1)$$

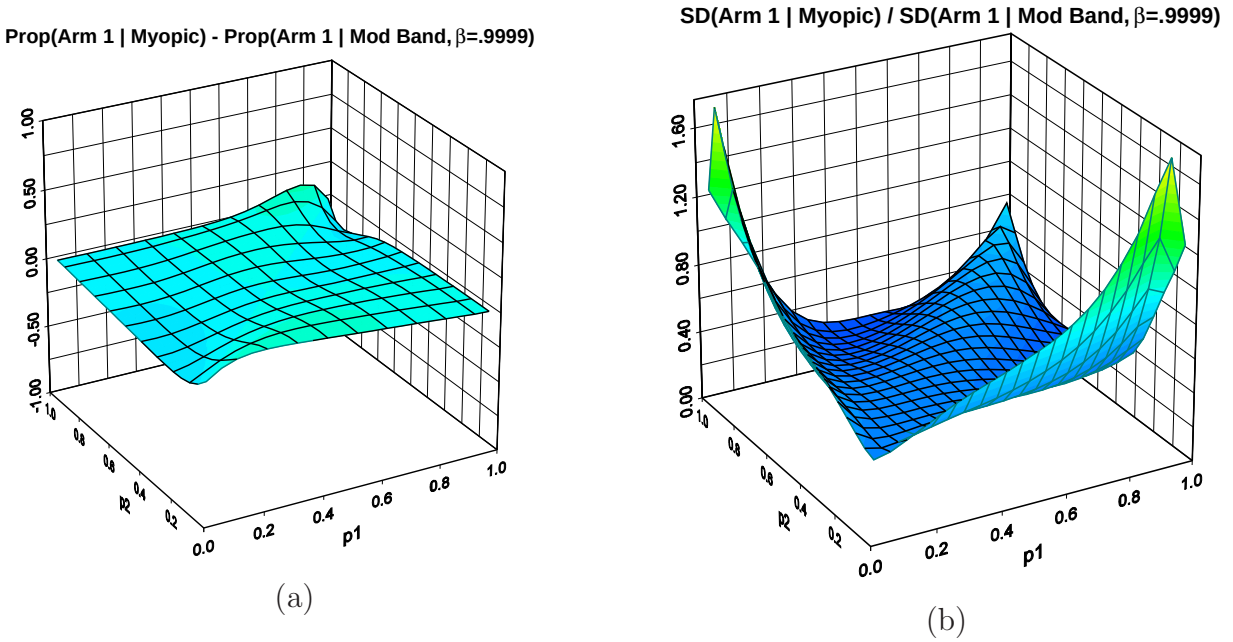
In the figure, the prior for each curve is indicated by "$(a_1, b_1)(a_2, b_2)$" situated close to the curve itself. While the data somewhat overlap, different line styles have been used to connect the data from each prior together. When the priors are $p_1 \sim Be(1, 2)$, $p_2 \sim Be(2, 1)$, all of the MB designs have $\mathcal{S}$ values very close to those of OS and MY. Also, the range of values for both $\mathcal{D}$ and $\mathcal{S}$ are much greater than for the uniform case. However, the arc shape and the ordering of the designs is quite similar to the uniform case. When the priors are $p_1 \sim Be(2, 1)$, $p_2 \sim Be(2, 1)$, the curve and the ordering are again similar to the uniform case. Further, the sampling efficiencies of all the MB designs are still very close to those of OS and MY. Not surprisingly, however, both the designs EA and RPW have improved $\mathcal{S}$ compared to the uniform case. In the third scenario, when $p_1 \sim Be(8, 8)$, $p_2 \sim Be(1, 1)$, the shape of the data appears somewhat different. There is a sharp decline in $\mathcal{D}$ values for the MB designs whereas, for the other prior configurations, MB8 in particular had very good decision efficiency. This is because the strong prior on the first arm has the effect that a few initial failures on the second arm will cause these designs to avoid further exploration of that arm, with the converse occurring for a few initial successes. Also, here, MB1 is essentially identical to OS on both measures, while for the priors with more equal strength, it typically had a smaller $\mathcal{D}$ value.

Pointwise Behavior:

To better understand the behavior of ad hoc designs and also to allow for better frequentist interpretations of the data, we examined pointwise plots of operating characteristics of (p_1, p_2) in the unit square. These include both Bayesian and frequentist versions of $\mathcal{S}$ and $\mathcal{D}$ along with the expected values and standard deviations of the number allocated to each arm. Naturally, with uniform priors, the values for the two arms are symmetric. We also reviewed pointwise data on the proportion of time each arm was determined to be best, the probability of correct selection, and the expected number of successes. Due to space considerations, we only discuss a couple of these characteristics.

One interesting feature of these data is that, with uniform priors, while for each (p_1, p_2) pair the MB4 and MY designs allocate nearly the same number of observations to Arm 1, the decision efficiency of these two designs is dramatically different. Figure 3(a) shows the expected difference in proportions allocated to Arm 1 for the two designs when $N = 50$. On average this difference is zero because each allocates $N/2$ to each arm. Still, even the interquartile range of this difference is small at (-3%, 3%) with maximum and minimum values of $\pm$ 15% occurring in the extreme regions close to (0,0) and (1,1). In trying to understand how the two designs could allocate so similarly yet differ so much on $\mathcal{D}$, we found that the underlying problem seemed to be that MY had a much higher standard deviation for those allocated to Arm 1.

Figure 3(b) displays the ratio of SD(# on Arm 1) for the MB4 versus MY. On average, MB4 is half as variable as MY. The standard deviation for MY is especially high near the diagonal because, after the initial mandatory observation on each arm, the design can easily begin choosing only one of the arms when it observes an early success on that arm. This will happen about half the time with each arm, making the variability extremely high when the success rates are close together. Further from the diagonal, MY can much more easily discern the correct arm to select quite early on. On the other hand, MB4 obtains much more information about each success rate because it thinks the trial will continue, in a sense,



(a) (b)

Figure 3 $N = 50$, uniform priors

forever, and that it therefore has plenty of time to keep observing arms without losing too much immediate gain. Since $\mathcal{D}$ is best when it has equal information on each arm, the MB4 makes better terminal decisions while maintaining roughly the same average allocations to the arms as does MY. These same behaviors of the two designs explain another interesting feature of Figure 3(b). Note the high relative variability of MB4 design near the extremes of (0,1) and (1,0). Here MY almost immediately ascertains the better arm and sticks to it and thus has very little variability. MB4, however, insists on continuing to sample "some" from each arm to satisfy the need to continue to explore. Thus MB4's variability is much higher than that for MY. In this case, the *more* variable sampling in this region also contributes to higher $\mathcal{D}$.

It is important to have access to a variety of pointwise operating characteristics since investigators need such detailed information about designs. In the past this type of exact distributional behavior of arbitrary operating characteristics has been unavailable, although such behaviors are occasionally approximated via simulation.

6 Discussion

In most experimental situations, there are many desirable but conflicting design considerations. Often, however, designs are either optimized with respect to a single objective or arbitrary tradeoffs among objectives are built in. Knowledge of optimal tradeoffs provides a point of reference from which to discuss those achieved by any given ad hoc design. Ad hoc designs can often be more desirable since they may be more intuitive or accepted among relevant investigators. As a result, analyses such as those carried out here can suggest appropriate ad hoc designs directed at the desired tradeoffs.

Here we have analyzed tradeoffs for two specific objectives. These are assessed via efficiency measures of both optimal patient care during the experiment and good decision making at its termination. In this case, it was found that, not only were there designs on the optimal tradeoff curve that very nearly optimized both criteria simultaneously, but also that there was a family of simple ad hoc designs, namely the modified bandit designs, that performed nearly as well.

Without a method for optimizing combinations of objectives, it would be impossible to fully evaluate such design complexities. Fortunately, by using a Bayesian approach, optimization can be achieved with dynamic programming. The resulting optimal design, and any ad hoc designs, can then be evaluated using path induction on a wide variety of operating characteristics including frequentist assessments such as pointwise behavior.

The technique used to optimize such tradeoffs is quite general, in that one can combine the objectives into a single one, with a parameter (α) to weigh the relative concerns. This can be extended to combinations of several objectives as well. For example, in the present example, one might wish to include expected sample size since this would allow for the incorporation of optional stopping. Various weighting schemes or "costs" would not change the approach. One could also take the more standard approach of specifying a required efficiency for one objective and determining the best efficiency available for the other. In this case α could be viewed as a control parameter, where a search is used to find the value of α that yields the required efficiency.

Another advantage to the approach used here is that it can aid in the development of good ad hoc designs when the necessary sample sizes are too large to optimize via dynamic programming. If a design makes very good tradeoffs on moderate sample sizes, it is likely the case that it will do well with larger sample sizes. One can also simulate properties of the ad hoc procedure since, at each state, the arm to be sampled can be determined quickly. What is lost is that, for the large sample size, one won't know exactly how close the design is to making optimal tradeoffs.

In conclusion, we have shown here that, for the commonly considered problem of simultaneously trying to optimize patient care and efficient decision making, there are two excellent classes of designs. The first and best of these are the designs that literally optimize the tradeoffs between the two defined objectives. The second are the modified bandit designs that lie extremely close to the optimal designs that weigh both objectives equally.

References

[1] Atkinson, A.C. and Donev, A.N. (1992), *Optimum Experimental Designs.* Oxford Statistical Sciences Series **8**, Oxford University Press, New York.

[2] Bai, Z., Hu, F. and Rosenberger, W.F. (2002), "Asymptotic properties of adaptive designs for clinical trials with delayed response", *Annals of Statistics* **30**: 122–139.

[3] Banks, J., Porter, D. and Olson, M. (1997), "An experimental analysis of the bandit problem", *Economic Theory* **10**: 55–77.

[4] Bather, J.A. (1985), "On the allocation of treatments in sequential medical trials", *Int.'l Stat. Review* **53**: 1–13.

[5] Bellman, R. (1957), *Dynamic Programming*. Princeton.

[6] Bergemann, D. and Valimaki, J. (2001), "Stationary multi-choice bandit problems", *Journal of Economic Dynamics and Control* **25**: 1585–1594.

[7] Biswas, A. (2003), "Generalized delayed response in randomized play-the-winner rule", *Comm. Stat. — Sim. and Comp.* **32**: 259–274.

[8] Bradt, R. and Karlin, S. (1956), "On the design and comparison of certain dichotomous experiments", *Ann. Math. Statist.* **27**: 390–409.

[9] Coad, D.S. and Ivanova, A. (2005), "Sequential urn designs with elimination for comparing $K \geq 3$ treatments", *Statistics in Medicine* **24**: 1995–2009.

[10] Fedorov, V.V. (1972), *Theory of Optimal Experiments.* Academic Press, New York.

[11] Ferguson, T. (1998), *Optimal Stopping and Applications*, www.math.ucla.edu/˜tom/Stopping/Contents.html.

[12] Ferguson, T. and Hardwick, J. (1988), "Stopping rules for proofreading", *J. App. Prob.* **26**: 304–313.

[13] Flournoy, N., Durham, S.D. and Rosenberger, W.F. (1995), "Toxicity in sequential dose-response experiments", *Sequential Analysis* **14**: 217–228.

[14] Fuh, C. and Hu, I. (2000), "Asymptotically efficient strategies for a stochastic scheduling problem with order constraints", *Anal. Statist.* **28**: 1670–1695.

[15] Gittins, J.C. and Jones, D.M. (1974), "A dynamic allocation index for the sequential design of experiments", *Progress in Statistics*, J. Gani et al., ed.'s, North Holland, 241–266.

[16] Gooley, T.A., Martin P.J., Fisher L.D. and Pettinger, M. (1994), "Simulation as a design tool for phase I/II clinical trials — an example from bone-marrow transplantation", *Controlled Clinical Trials* **15**: 450–462.

[17] Hardwick, J.P. (1989), "Recent progress in clinical trials that adapt for ethical purposes", *Statist. Sci.* **4**: 328–336.

[18] Hardwick, J. (1995), "A modified bandit as an approach to ethical allocation in clinical trials", *Adaptive Designs: Institute Math. Stat. Lecture Notes* **25**, B. Rosenberger & N. Flournoy, ed.'s, 65–87.

[19] Hardwick, J., Oehmke, R., and Stout, Q.F. (2006), "New adaptive designs for delayed response models", *J. Stat. Plan. and Infer.* **136**: 1940–1955.

[20] Hardwick, J. and Stout, Q.F. (1996), "Optimal allocation for estimating the mean of a bivariate polynomial", *Sequential Analysis* **15**: 71–9

[21] Hardwick, J. and Stout, Q.F. (1999), "Path induction for evaluating sequential allocation procedures", *SIAM J. Scientific Computing* **21**: 67–87.

[22] Hardwick, J. and Stout, Q.F. (2002), "Optimal few-stage designs", *J. Stat. Plan. and Infer.* **104**: 121–145.

[23] Hardwick, J. and Stout, Q.F. (2005), "Response adaptive designs that incorporate switching costs and constraints", *Simulation 2005*, V.B. Melas, ed., NII Chemistry St. Petersburg, 305–312.

[24] Ivanova, A. and Rosenberger, W. (2000), "A comparison of urn designs for randomized clinical trials of $k > 2$ treatments", *J. Biopharm. Statist.* **10**: 93–107.

[25] Jones, P.W., Lewis, A.M. and Hartley, R. (1995) "Some designs for multicriteria bandits", *Adaptive Designs: Institute Math. Stat. Lecture Notes* **25**, B. Rosenberger & N. Flournoy, ed.'s, 88–94.

[26] Robbins, H. (1952) "Some aspects of the sequential design of experiments", *Bull. Amer. Math. Society* **55**: 527-535.

[27] Rosenberger, W. (1999), "Randomized play-the-winner clinical trials: review and recommendations", *Controlled Clinical Trials* **20**: 328–342.

[28] Simon, R. (1989), "Optimal two stage designs for phase II clinical trials", *Controlled Clinical Trials* **10**: 1–10.

[29] Stallard, N. and Rosenberger, W.F. (2002), "Exact group-sequential designs for clinical trials with randomized play-the-winner allocation", *Statistics in Medicine* **21**: 467–480.

[30] Stout, Q.F. and Hardwick, J. (2005), "Parallel programs for adaptive designs", *Handbook on Parallel Computing and Statistics*, E. Kontoghiorghes, ed., Marcel Dekker, 347–373.

[31] Thompson, W.R. (1933), "On the likelihood that one unknown population exceeds another in view of the evidence of the two samples", *Biometrika* **25**: 275–294.

[32] Wang, Y.-G. (1991), "Sequential allocation in clinical trials", *Comm. Stat. — Theor. Meth.* **20**: 791–805.

[33] Wang, Y.-G. (1991), "Gittins indices and constrained allocation in clinical trials", *Biometrika* **78**: 101–111.

[34] Wei, L.J. and Durham, S. (1978), "The randomized play-the-winner rule in medical trials", *J. Amer. Statist. Assoc.* **73**: 830–843.

[35] Wei, L.J. (1988), "Exact two-sample permutation tests based on the randomized play-the-winner rule", *Biometrika* **75**: 603–606.

[36] Woodroofe, M. (1982), "Sequential allocation with covariates", *Sankhya* **44**: 403–414.