Journal of Sequential Planning and Inference (2006), vol. 136, pp. 1940–1955.

New Adaptive Designs for Delayed Response Models

Janis Hardwick Robert Oehmke Quentin F. Stout
University of Michigan
Ann Arbor, Michigan 48109 USA

Abstract

Adaptive designs are effective mechanisms for flexibly allocating experimental resources. In clinical trials particularly, such designs allow researchers to balance short and long term goals. Unfortunately, *fully* sequential strategies require outcomes from all previous allocations prior to the next allocation. This can prolong an experiment unduly. As a result, we seek designs for models that specifically incorporate delays.

We utilize a delay model in which patients arrive according to a Poisson process and their response times are exponential. We examine three designs with an eye towards minimizing patient losses: a delayed two armed bandit rule which is optimal for the model and objective of interest; a newly proposed hyperopic rule; and a randomized play-the-winner rule. The results show that, except when the delay rate is several orders of magnitude different than the patient arrival rate, the delayed response bandit is nearly as efficient as the immediate response bandit. The delayed hyperopic design also performs extremely well throughout the range of delays, despite the fact that the rate of delay is not one of its design parameters. The delayed randomized play-the-winner rule is far less efficient than either of the other methods.

Keywords and phrases: optimal allocation, two-arm bandit, sequential sampling, design of experiments, clinical trial, dynamic programming, hyperopic

1 Introduction

Adaptive or sequential designs take advantage of accruing information to optimize experimental objectives. Such designs have long been proposed as models for clinical trials. While the primary goal for a trial may be to evaluate treatment options with the intention of improving treatment for patients who come after the experiment, the well-being of the patients within the study is also an important consideration. Adaptive designs address this tradeoff far better than do classical fixed allocation designs, and a good adaptive design requires fewer experimental resources (including patients) to achieve the same statistical goals as a fixed design.

Unfortunately, adaptive designs often possess features that inhibit their use. In particular, the ethical questions posed by adaptive versus fixed sample designs are controversial and complex. Certainly, there is no ethically "correct" viewpoint. However, we favor the idea of providing investigators with flexible options so that they can choose a statistical design appropriate to their own setting.

Another difficulty with adaptive methods is that statistical analyses of data arising from such designs is generally more complicated than it is for fixed designs. Exact solutions, in particular, command extreme

computer resources and complex algorithms which have only recently begun to be available. We tackle that problem here with refined parallel algorithms detailed in Hardwick et al. (1999) and Oehmke et al. (2001).

A third concern is the fact that fully sequential designs require all earlier responses to be in hand before allocating the next patient. Short of staged designs, which can diminish the impact of delayed responses, there is little in the literature relating to this problem. Specifically, we know of no non-trivial models incorporating delayed responses for which exactly optimal adaptive designs have been obtained.

In this paper, we seek to optimize an objective function for a problem in which there are two populations. The responses, which may be delayed, are independent Bernoulli random variables. Patient response times follow independent exponential distributions depending on their treatment assignment. We impose a delay structure in which patients arrive via a Poisson process. We assume that the arrival rate and the mean response times are known.

In our examples, the objective function is to maximize expected patient successes during the experiment. This expectation is taken with respect to a Bayesian model with independent beta priors on the success probabilities of the two arms. Given this, it is reasonable to model the problem as a 2-armed bandit (2AB) with delayed response. Recall that the objective of a bandit problem is to allocate resources to different experimental "arms" in such a way that the total return from the experiment is optimized. In this case, the return or objective corresponds to patient successes.

There has been some work done on the related problem of maximizing patient survival times in a 1-armed bandit (1AB) model. In the 1AB there are actually two arms, but the attributes of one of them are completely known. Eick (1988) addresses the extent to which geometric response delays affect standard behavioral characteristics of the 1AB, where the survival rate of one arm is known and the goal is to maximize total survival time by allocating patients to either the "known" or unknown therapy. Some of these results have been extended and generalized by Wang (2000). Tantiyaswasdikul (1992) examines a 1AB with single covariate model where the response may be delayed up to M stages, with a different known probability of incurring a delay of each length m, $m = 1, \ldots, M$.

There are a couple of other relevant scenarios in which delayed response designs have been studied for binomial populations. In each case, the ethical goal of optimizing patient successes is a key design consideration. Primarily, the focus has been on urn models. Particularly popular are delayed response variations of the randomized play-the-winner rule (RPW) of Wei and Durham (1978). See Section 3 for details. Douke (1994) and Langenberg and Srinivasan (1982) have shown interest in delayed response versions of the well known two-stage design of Colton (1963). The newly proposed hyperopic design described in Section 4 also utilizes two-stage concepts.

In the next section we develop models for the delayed response bandit and present the requisite dynamic programming equations. One natural dynamic programming approach is computationally difficult. This is described in Appendix A. A second approach, which is more amenable to computer implementation, is outlined in Appendix B. In Sections 3 and 4 we describe a delayed RPW rule and a delayed hyperopic rule, respectively. In Section 5 we compare the delayed versions of the three rules not only with each other but also with the non-delayed versions of each. In the last section, Section 6, we discuss our findings.

2 Models with Exponential Delay

Suppose that patients arrive according to a Poisson process with rate λ_s. As they arrive, they are assigned either to arm (treatment) 1 or 2. Patient responses are Bernoulli with success rates π_1 and π_2. Prior

distributions on the π_i are Be(a_i, b_i), $i = 1, 2$, respectively. The response time for a patient on arm i is exponential with mean $1/\lambda_i$, $i = 1, 2$. Response times are independent themselves and independent of arrival times and of actual responses. The experiment will allocate a total of n patients. For an arm i, we use $\neg\, i$ to denote the other arm, i.e., arm $3-i$.

If a patient arrival occurs at time t, the patient is allocated to arm 1 or 2 based on data collected up until t. This includes the responses and number of patients allocated to each arm, as well as the priors. A sufficient statistic at time t is $\langle s_1(t), f_1(t), u_1(t); s_2(t), f_2(t), u_2(t)\rangle$, where $s_i(t)$, $f_i(t)$ are the number of successes and failures on arm i and $u_i(t)$ is the number outstanding on arm i, $i = 1, 2$. Because the problem is stationary in time, we can drop the t notation. Thus an allocation policy is a function that depends on the priors and n and maps $\langle s_1, f_1, u_1; s_2, f_2, u_2\rangle$ to $\{1, 2\}$. Optimal solutions are policies that are optimized for a given objective function. Since the objective here is to maximize total experimental patient successes, this problem has the form of a two armed bandit with delay. We call this optimization problem the *delayed 2-armed bandit*, D2AB. Note, however, that our approach also applies to numerous other objective functions.

It is well-known that such optimization problems can be solved via dynamic programming. However, computational space and time grow exponentially in the number of arms, and the delay complicates this further. The state space involves all possible variations of its components, as long as all are nonnegative and their sum is no greater than n. There are thus $\binom{n+6}{6} = \Theta(n^6)$ states in the D2AB. More generally, the delayed k-arm bandit will have $\binom{n+3k}{3k}$ states. This is in contrast to the $\Theta(n^4)$ states in the standard 2AB, and the general case of $\binom{n+2k}{2k}$ states in the standard k-arm Bernoulli bandit. In fact, the states in the D2AB are in a natural 1-1 correspondence with the states in the standard 3AB. Here we will concentrate on the D2AB, although with simple changes the techniques could be applied to general delayed k-arm bandits. However, such solutions are currently not computationally practical when $k > 2$ and sample sizes are > 100.

To apply dynamic programming, one needs to know the terminal states, i.e., those states which can be directly evaluated without recourse to recursion. In this situation, it is those states for which $u_1 = u_2 = 0$ and $s_1 + f_1 + s_2 + f_2 = n$; i.e., those states for which all n patients have been allocated and all of their responses have been observed. For our primary example of trying to maximize successes, the *value* of a terminal state is simply $s_1 + s_2$. Ultimately, our goal is to determine the value, V, of the initial state $\langle 0, 0, 0; 0, 0, 0\rangle$.

There are various ways to tackle this problem, and finding one that is computationally feasible is a keystone of the solution. We consider two alternatives — one suited to describing characteristics of the solution and the other to solving the problem computationally. In each case we develop recursive dynamic programming equations. Let $\pi_i(s_i, f_i)$ denote the posterior probability that an observation on arm i will be a success, given that s_i successes and f_i failures have been observed previously on the arm. Thus, $\pi_i(s_i, f_i) \sim \mathrm{Be}(a_i+s_i, b_i+f_i)$ for $1 = 1, 2$.

Approach I: Perhaps the most natural approach is the one in which time is marked by patient arrivals, because these are the only times when action is taken and decisions are needed. In the meantime, outstanding responses may come in, possibly including that of the patient most recently assigned. For the u_i patients with unobserved outcomes on arm i, the number of newly observed successes, s_i', or failures, f_i', before the next patient arrives must satisfy $0 \le s_i'$, $0 \le f_i'$, and $s_i' + f_i' \le u_i$. Thus, by the time the next patient arrives, the system may have moved to any of $\binom{u_1+2}{2}\binom{u_2+2}{2}$ different states.

The dynamic programming equations for this approach are presented in Appendix A. Unfortunately, evaluating these equations requires $\Theta(n^{10})$ time, and thus they are computationally infeasible except for trivial sample sizes.

Approach II: A second approach marks time by *events*, where an event is either a subject arrival or a response from one of the arms. Because we are using continuous time, we can assume that only one event occurs at a time. Let $P_1(u_1, u_2)$, $P_2(u_1, u_2)$, $P_s(u_1, u_2)$ represent the probability that the next event is an observation on arm 1, an observation on arm 2, or a subject arrival, respectively. While P_1, P_2 and P_s are interrelated, they have a simple form, namely

$$P_s(u_1, u_2) = \frac{\lambda_s}{\lambda_s + u_1 \cdot \lambda_1 + u_2 \cdot \lambda_2} \quad \text{and} \quad P_i(u_1, u_2) = \frac{u_i \cdot \lambda_i}{\lambda_s + u_1 \cdot \lambda_1 + u_2 \cdot \lambda_2}.$$

Let $\sigma + \widehat{y}$ denote state σ with component y increased by one. Then, the dynamic programming equation is as follows.

$$\begin{aligned} V(\sigma) \;=\; & P_1(u_1, u_2) * \left[\pi_1(s_1, f_1) \cdot V(\sigma + \widehat{s_1} - \widehat{u_1}) + (1 - \pi_1(s_1, f_1)) \cdot V(\sigma + \widehat{f_1} - \widehat{u_1})\right] \\ & + P_2(u_1, u_2) * \left[\pi_2(s_2, f_2) \cdot V(\sigma + \widehat{s_2} - \widehat{u_2}) + (1 - \pi_2(s_2, f_2)) \cdot V(\sigma + \widehat{f_2} - \widehat{u_2})\right] \\ & + P_s(u_1, u_2) * \max\left\{V(\sigma + \widehat{u_1}),\ V(\sigma + \widehat{u_2})\right\} \end{aligned}$$

Here, the allocation choice is handled in the last term, where if there is a subject arrival then we just determine to which arm we allocate. Initially this simply means that the arm has one more unobserved allocation. The advantage of this approach is that each state depends upon only 6 others, rather than the $O(n^4)$ of Approach I, so the computations can be completed in $\Theta(n^6)$ time. While still formidable, this can be achieved for useful sample sizes, as explained in Appendix B and in Oehmke et al. (2001).

Note that when the sample size has been reached then the third term of the recurrence is eliminated, and the formulae for P_1 and P_2 are adjusted so that $P_i(u_1, u_2) = u_i\lambda_i/(u_1\lambda_1 + u_2\lambda_2)$. Similarly, when σ is a terminal state, i.e., when all observations have been obtained ($s_1 + f_1 + s_2 + f_2 = n$), then $V(\sigma)$ is the value of the objective function being optimized. In this paper, $V(\sigma) = s_1 + s_2$.

Exact evaluations of arbitrary, sub-optimal allocation designs for the present problem are possible via slight modifications to the algorithm in Appendix B. This is accomplished by replacing the $\max\{V(\sigma + \widehat{u_1}), V(\sigma + \widehat{u_2})\}$ in the recurrence for $V(\sigma)$ with the allocation decision that the design would make. Note that this decision may be stochastic.

3 A Randomized Play-the-Winner Rule

One popular ad hoc adaptive rule is the randomized play the winner (RPW) rule which first appeared in Wei and Durham (1978), and which has subsequently been utilized in a number of clinical trials. In this urn model, there are initial balls representing the treatment options. These may be thought of as a prior on success rates π_1 and π_2. One starts with an urn containing α_i balls of type i for $i = 1, 2$. Patients are assigned to arms according to the type of ball drawn at random from the urn. Sampling is with replacement, and balls are added to the urn according to the last patient's response. If the patient 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 of type $\neg\, i$ are added to the urn. Most often, $\alpha_1 = \alpha_2$ and $\beta_s = \beta_f = 1$.

One advantage of urn models like RPW is the natural way in which delayed observations can be incorporated into the allocation process. When a delayed response eventually comes in, balls of the appropriate type are added to the urn. Since sampling is with replacement, any delay pattern can be accommodated. We call this design the *delayed RPW* rule (DRPW).

Much of the literature on adaptive designs that incorporate delayed observations use the DRPW as the allocation procedure. In Bandyopadhyay and Biswas (1996), the authors consider a slightly altered version of this rule for a related best selection problem. Biswas (2003) utilizes regression delay structures and examines the DRPW's performance. Rosenberger (1999) considers the DRPW in a discussion paper on the RPW. There, one relevant remark is that "Exact computations preclude delayed response because of theoretical difficulties, ...". Exact evaluations of the DRPW appear in Hardwick et al. (2001), in which the present exponential delay structure is imposed. Bai et al. (2002) present asymptotic results relating to urn composition and the limiting distribution of estimators for a DRPW design.

4 Adaptive Hyperopic Designs

One way to approach delayed response problems is to utilize *hyperopic* designs, where a hyperopic design is one which makes sampling decisions based on the trial length, in addition to the priors. For example, most 1-stage and 2-stage designs are of this form, in both frequentist and Bayesian frameworks. This is in contrast to myopic designs where optimization decisions assume only one more observation will occur. Note that RPW can be viewed as being myopic, while the optimal solution of the 2AB is hyperopic.

We create an *adaptive hyperopic* design, H, as follows:

1. Start with a class $\mathcal{D}$ of (simple) hyperopic designs for immediate responses. For example, one might use standard 1-stage designs, which use the best fixed sample size rule for allocation.

2. As the experiment proceeds, at each state $\sigma = \langle s_1,f_1,u_1; s_2,f_2,u_2\rangle$, with m remaining observations to be assigned, determine the optimal fixed sample design $D_\sigma \in \mathcal{D}$ for an experiment having $m + u_1 + u_2$ observations, subject to the condition that at least u_1 observations must be made on arm 1 and u_2 must be made on arm 2.

3. If D_σ would initially allocate more new observations to arm i than to arm $\neg\, i$, then H(σ) would allocate the next observation to arm i; while if D_σ would allocate equal numbers of new observations then H(σ) would randomize the next allocation. In some settings one might prefer to increase randomization by allocating the next observation to arm i with probability proportional to the number of new observations D_σ would allocate to it.

When the objective is to maximize successes, the optimal 1-stage design allocates all observations to the arm with the highest prior mean; that is, it makes the myopic choice. To encourage exploration, for $\mathcal{D}$ we instead use a simplistic 2-stage design that is invoked at each state σ as the experiment proceeds: let $W_i(\sigma)$ be the value of the 2-stage immediate response design in which the first stage allocates $1 + u_i$ observations to arm i and $u_{\neg\, i}$ observations to arm $\neg\, i$. In the second stage, it allocates the remaining $n - 1 - u_1 - s_1 - f_1 - u_2 - s_2 - f_2$ observations to the arm with the highest posterior mean. If $W_i(\sigma) > W_{\neg\, i}(\sigma)$ then we allocate the next observation to arm i, while if they are equal then we randomize between the arms. We use DAH to denote this adaptive hyperopic design for delayed responses. Note that if one arm has no pending observations while the other has many then DAH will tend to sample the former.

One can prove that if $\mathcal{D}$ is the class of optimal 2-stage designs then the corresponding adaptive hyperopic design is asymptotically optimal, and that it is not asymptotically optimal for the restricted 2-stage designs used here. However, despite the simplicity of the 2-stage hyperopic design used at each state, the sequential updating of the information results in a nearly optimal design, even in the presence of delays, as is shown in Section 5. Restricting $\mathcal{D}$ to 2-stage designs that add only one observation to the first stage (rather than considering all possible additions) greatly reduces the computations required at each stage,

making it more likely to be applied in practice. Simulations can be easily used to produce approximate evaluations of its expected value and operating characteristics.

Exact evaluations, however, are as complex as those for the D2AB since there are still $\Theta(n^6)$ states. Note that attaining this time is somewhat complicated. A 2-stage design for immediate responses which allocates an initial o_i observations on arm i has $(o_1+1)(o_2+1)$ outcomes at the end of the first stage. Hence an exact evaluation of an adaptive hyperopic design which used a straightforward examination of all possible 2-stage allocations at each state would take $\Theta(n^{10})$ total time. Even by restricting to only an examination of the 2 simplest allocations at each state, as is done here, would take $\Theta(n^8)$ time in a naive approach. This can be reduced to $\Theta(n^6)$, but the steps needed to do this are beyond the scope of this paper and will be published elsewhere.

5 Results of Comparisons

We have carried out exact analyses of the exponential delay model for D2AB, DRPW, and DAH. In these preliminary analyses we take $n = 100$, and for the DRPW take $\alpha_1 = \alpha_2 = \beta_s = \beta_f = 1$. To simplify comparisons, throughout we take λ_s, the patient arrival rate, to be 1, and vary the response rates. Since it is relative rates which determine the behavior, when $\lambda_1 = \lambda_2 = \lambda$ one could set $\lambda_s = 1/\lambda$ and view the results as fixing the response rate at 1 and varying the arrival rate.

We look first at base case and best case scenarios. For comparative purposes, we take the best fixed in advance allocation procedure to be the base case, i.e., the optimal solution when no responses will be available until after all n patients have been allocated. To maximize expected successes one should allocate all patients to the treatment with the higher expected success rate. We denote the expected number of successes in the base case by $\mathrm{E}_{base}[\mathrm{S}]$. Throughout we only report expected values, not variances nor other distributional information, because the overwhelming source of variation is the uncertainty built into the priors.

Results for uniform priors: We initially consider uniform priors on the treatment success rates π_1 and π_2, in which case all fixed allocations result in the same expected successes. For these priors, $\mathrm{E}_{base}[\mathrm{S}] = n/2$.

We encounter the best possible case when all responses are observed immediately (full information). In this situation, DRPW is simply the regular RPW and the D2AB is the regular 2-armed bandit. DAH becomes an adaptive hyperopic design which does not seem to have been previously analyzed, and in future work we will examine adaptive hyperopic designs for various optimization problems. Recall that the regular 2-armed bandit optimizes the problem of allocating to maximize total successes. Letting $\mathrm{E}_{opt}[\mathrm{S}]$ represent expected successes in the best case (i.e., those obtained by the regular 2-armed bandit), for our example we have $\mathrm{E}_{opt}[\mathrm{S}] = 64.9$. Using the difference $\mathrm{E}_{opt}[\mathrm{S}] - \mathrm{E}_{base}[\mathrm{S}]$ as a scale for improvement, we can think of the values on this scale, (0, 14.9), as representing the "extra" successes over the best fixed allocation of 100 observations. We take $\mathrm{R}(\delta) = (\mathrm{E}_{\delta}[\mathrm{S}] - \mathrm{E}_{base}[\mathrm{S}]) / (\mathrm{E}_{opt}[\mathrm{S}] - \mathrm{E}_{base}[\mathrm{S}])$ to be the *relative improvement* over the base case for any allocation rule δ. Note that $\mathrm{R}(\delta)$ also depends on n and the prior parameters.

For fixed priors, arrival rate, and response rates, one can show that as $n \to \infty$, R(D2AB) $\to 1$, R(DAH) $\to C_{\mathrm{DAH}}$, and R(DRPW) $\to C_{\mathrm{DRPW}}$, where C_{DAH} and C_{DRPW} are constants less than one. However, this asymptotic behavior gives little information about the values for practical sample sizes, and hence one needs to determine their behavior computationally. As will be seen, the asymptotic suboptimality of DAH is particularly misleading.

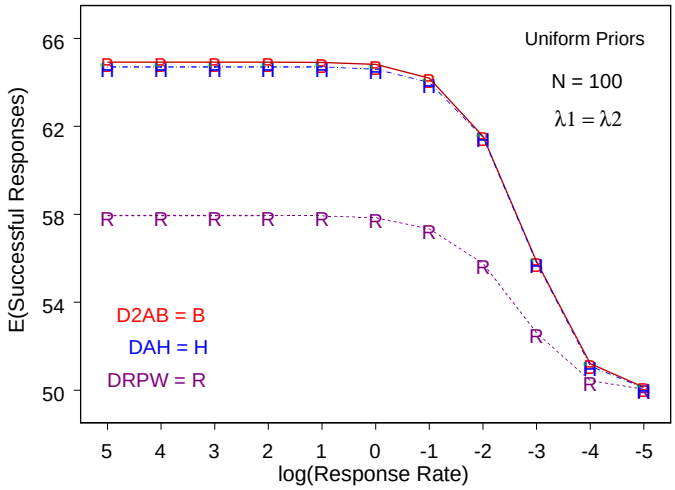

Figure 1: Expected successes for D2AB, DRPW, and DAH, $\lambda_s = 1$, $\lambda_1 = \lambda_2$

Tables 1, 2 and 3 contain the expected successes for the D2AB, DRPW, and DAH rules, respectively. Patient response rates, λ_1 and λ_2, vary over a grid of values between 10^{-5} and 10^{1} and the patient arrival rate, λ_s, is fixed at 1. Note that, for all rules, when $\lambda_1 = \lambda_2 = 10^{-5}$, E[S] $\approx$ 50. When $\lambda_1 = \lambda_2 \geq 10$ all of the rules achieve a number of successes nearly identical to the number they achieve in the immediate response situation. Under these conditions, the D2AB rule achieves E[S]=64.9, and the DAH rule achieves E[S]=64.7. Note that for the DRPW, E[S] = 57.9, which gives an R of 0.53. With the immediate response RPW, we can expect to gain only 7.9 successes as compared to the 14.9 obtained by the optimal 2AB and the 14.7 for the hyperopic design.

Moving away from the extreme points, consider the case when $\lambda_1 = \lambda_2 = 10^{-1}$, one order of magnitude smaller than λ_s. All of the rules can be seen to be quite robust to such delays, even though asymptotically there are, on average, nearly 10 patients allocated but unobserved. It is only when *both* response rates are at least three orders of magnitude below the arrival rate that results begin to degrade seriously. When $\lambda_1 = \lambda_2 = 10^{-3}$, for example, the relative improvement over the base case of the D2AB and the DAH is only 0.40, and for the DRPW it is a dismal 0.17. It is also interesting to note that even when the response rate is only 1/100$^{\text{th}}$ the arrival rate, the D2AB and the DAH do better than the RPW with immediate responses. Figure 1 illustrates the relative performance of the rules when the arrival rate is one and the response rates vary between 10^{-5} and 10^{5}.

One conspicuous feature of this figure is the fact that the DAH is nearly as good as the D2AB throughout the entire range. This is particularly interesting because the D2AB has the delay as one of its design parameters while the DAH does not. Hence one can view the graph as comparing the expected values of a family of bandit designs, one per delay, to the operational characteristics of a single hyperopic design.

When we consider scenarios in which only one treatment arm supplies information to the system, we see interesting behavior. For example, when $\lambda_1 = \lambda_s = 1$ but $\lambda_2 = 10^{-5}$, the relative improvement is 0.76 for the D2AB, 0.71 for the DAH, and 0.47 for the DRPW. The D2AB is now clearly superior to the DAH and the DRPW. However, the DRPW is interesting here since its R-value is 89% of that for the undelayed RPW value. Thus, since the RPW rule starts out so poorly, it has relatively less to lose. Note that with the D2AB one only loses 24% of the optimal solution while excluding half the information.

λ_1	λ_2						
$\downarrow$	10^{-5}	10^{-4}	10^{-3}	10^{-2}	10^{-1}	10^{0}	10^{1}
10^{-5}	50.1						
10^{-4}	51.2	51.2					
10^{-3}	55.4	55.4	55.8				
10^{-2}	59.3	59.4	59.9	61.5			
10^{-1}	60.9	61.0	61.6	63.1	64.1		
10^{0}	61.3	61.3	61.9	63.5	64.5	64.8	
10^{1}	61.3	61.3	62.0	63.5	64.6	64.8	64.9

Table 1: D2AB: E[S] as (λ_1, λ_2) vary, $n = 100$, $\lambda_s = 1$, uniform priors

λ_1	λ_2						
$\downarrow$	10^{-5}	10^{-4}	10^{-3}	10^{-2}	10^{-1}	10^{0}	10^{1}
10^{-5}	50.0						
10^{-4}	50.2	50.4					
10^{-3}	51.6	51.7	52.6				
10^{-2}	54.8	54.8	54.9	55.7			
10^{-1}	56.5	56.5	56.5	56.7	57.3		
10^{0}	56.9	56.9	56.9	57.1	57.6	57.8	
10^{1}	57.0	57.0	57.0	57.2	57.6	57.8	57.9

Table 2: DRPW: E[S] as (λ_1, λ_2) vary, $n = 100$, $\lambda_s = 1$, uniform priors

λ_1	λ_2						
$\downarrow$	10^{-5}	10^{-4}	10^{-3}	10^{-2}	10^{-1}	10^{0}	10^{1}
10^{-5}	50.1						
10^{-4}	50.7	51.2					
10^{-3}	53.8	54.1	55.8				
10^{-2}	58.4	58.5	59.4	61.5			
10^{-1}	60.3	60.4	61.2	62.9	64.0		
10^{0}	60.6	60.7	61.6	63.3	64.4	64.7	
10^{1}	60.6	60.8	61.7	63.3	64.4	64.7	64.7

Table 3: DAH: E[S] as (λ_1, λ_2) vary, $n = 100$, $\lambda_s = 1$, uniform priors

λ_1	λ_2						
$\downarrow$	10^{-5}	10^{-4}	10^{-3}	10^{-2}	10^{-1}	10^{0}	10^{1}
10^{-5}	50.0	50.0	50.2	53.4	55.3	55.7	55.7
10^{-4}	50.5	50.5	50.6	53.5	55.3	55.7	55.7
10^{-3}	52.6	52.7	52.9	54.4	56.1	56.5	56.5
10^{-2}	55.4	55.5	55.8	56.8	57.9	58.2	58.3
10^{-1}	56.7	56.7	57.1	58.2	59.1	59.3	59.3
10^{0}	56.9	57.0	57.3	58.5	59.4	59.6	59.6
10^{1}	56.9	57.0	57.3	58.5	59.4	59.6	59.7

Table 4: D2AB: E[S] as (λ_1, λ_2) vary, $n = 100$, $\lambda_s = 1$
Prior distributions are Be(1,1) and Be(1,1.5).

λ	Be(1,1)	Be(1,4)	Be(4,1)
10^{-0}	0.993	0.994	0.988
10^{-1}	0.952	0.946	0.929
10^{-2}	0.774	0.712	0.701
10^{-3}	0.393	0.304	0.309
10^{-4}	0.081	0.056	0.055

Table 5: Relative Efficiency for D2AB, $n = 100$, $\lambda = \lambda_1 = \lambda_2$, $\lambda_s = 1$
Both arms have the same prior distribution.

Some results for alternative priors: We include a couple of examples of how the performance of the D2AB changes when non-uniform priors are used. First we consider a case in which the priors for the two arms differ. In Table 4, prior distributions on π_1 and π_2 are Be(1,1) and Be(1,1.5), giving arms 1 and 2 prior means of 0.5 and 0.4, respectively. Note first that E_{opt}[S] = 59.6 which is down from 64.9 in the uniform case. This is a natural result of having one mean smaller than before. More interesting perhaps is the asymmetry in the tabulated values. The total expected successes are reduced more when the response rates for the inferior arm (as opposed to the better arm) are slow coming in.

In Table 5 we show the behavior of the D2AB as the response rate varies from rapid to extremely slow. Both arms have the same priors. Note that, while all priors exhibit the same basic behavior, there are noticeable differences in how well the D2AB does when the response rate is extremely slow, i.e., when there are very few observations during the trial.

Other differences show up when the non-delayed 2AB is compared to the omniscient design which always allocates to the better arm. For n=100 and Be(1,1) priors, the non-delayed 2AB attains 0.89 of the improvement that the omniscient design gets, for Be(1,4) it attains 0.80, and for Be(4,1) it attains 0.84. Combining this information with the fact that the values in Table 5 are relative to the non-delayed 2AB, not to the omniscient design, one sees that the D2AB does significantly better with uniform priors than with the others.

One way to view this problem independently from the allocation rules is to examine the expected number of allocated but unobserved patients when a new patient allocation decision must be made. Figure 2 displays this information for various delay rates relative to a patient arrival rate of $\lambda_s = 1$. In this figure, it is assumed that $\lambda_1 = \lambda_2 = \lambda$, and there are separate curves for $\lambda = 10^{-3}$, 10^{-2}, 10^{-1} and 1. As noted,

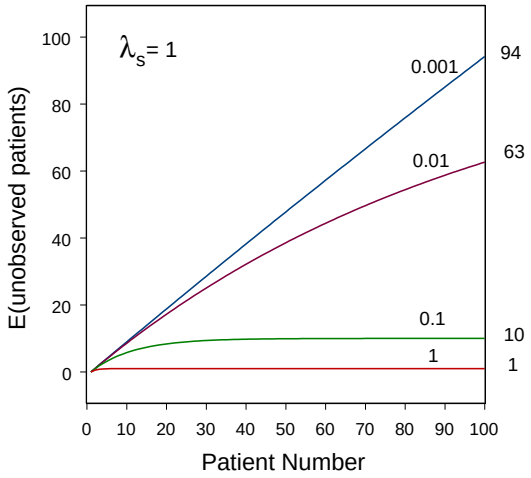

Figure 2: Expected number of unobserved patients when new patient arrives.
Curve label is $\lambda = \lambda_1 = \lambda_2$.

when the response delay rate is 1, at any point in time one expects only a single observation to be delayed, and the impact on performance is minimal. Once approximately 20 patients have been allocated, there is a consistent lag of about ten patients when $\lambda = 0.1$. Connecting this value to the results in Tables 1, 2, and 3, one finds that a loss of roughly 10% of the total information at the time of allocation of the last patient (and a significantly higher loss rate for earlier decisions), corresponds to a loss of only about 5% in terms of the improvement available from each rule.

When the response rate is about 100 times slower than the arrival rate, asymptotically there will be approximately 100 unobserved patients at any point in time. Fortunately, for a sample size of 100, one is quite far from this asymptotic behavior, and approximately 37% of the responses have been observed by the time the last allocation decision must be made. This allows the D2AB to achieve 77% of the relative improvement possible.

6 Discussion

Because there has been so little research addressing optimal adaptive designs with delayed responses, or addressing exact evaluations of general designs with delayed responses, there are numerous outstanding problems in the area. One might argue that exactly optimal designs aren't necessary in practice, especially if good ad hoc options are available. However, without a basis of comparison it is difficult to know how good ad hoc options are, since asymptotic analyses give only vague information about their behavior for practical sample sizes. Examining the properties of optimal designs can also lead to the development and selection of superior sub-optimal alternatives. In Section 4, for example, we propose ad hoc rules that come far closer to achieving optimal performance than does the DRPW rule, which is the rule most often suggested for delayed response scenarios. Still, to progress further much work needs to be done, particularly with regard to practical characteristics of these designs.

Two important concerns are a design's robustness and the ability to apply it flexibly. Due to space constraints such analyses could not be undertaken here, but previous analyses for fully sequential designs lead one to suspect that the D2AB optimized for one set of parameters and then evaluated with respect to

a second set will be nearly as efficient as the D2AB optimized for the second set. In future work we will examine the robustness and operating characteristics of the D2AB and DAH.

One way to address the model assumptions to improve robustness might be to use prior distributions on the response rate parameters. Also, it would also be extremely helpful to know the impact of the assumption of exponential response times. Unfortunately, optimizing and evaluating different arrival and response delay models can involve significantly different recursive equations, and the computational requirements can vary dramatically. For example, if response delays and arrivals occur at a constant rate, then the appropriate state space would be the successes and failures observed, coupled with the listing, in order, of the allocations not yet observed. This model has $\Theta(n^4 2^d)$ states, where d is the response delay expressed in terms of the arrival rate. For small to modest d, such as the $d = 2$ used in Rosenberger (1999), constant delay models can be optimized exactly, and arbitrary designs can be evaluated exactly, although we have not undertaken such an evaluation. For large d, however, this is currently infeasible. Certainly one can test the data's goodness of fit to the exponential model after the experiment, but this is not very helpful if the data fit it poorly. However, there is reason to believe that a variety of response time models would benefit from application of the optimal design for the exponential model. The resulting designs may not be optimal, but they will likely be very good.

Recall that one goal of this research is to develop exactly optimal, or nearly optimal, delayed response designs that allow for the use of *any* objective function, not just the bandit objective of maximizing reward (successes). The dynamic programming algorithm so developed in Appendix B, and its implementation on serial and parallel computers, has this capability; and in future work we will examine the performance of optimal designs for other objectives.

The adaptive hyperopic design approach is also extremely general, applicable to arbitrary objectives, and there are many variations possible for both immediate and delayed responses. An adaptive hyperopic design based on an asymptotically optimal family of designs is also asymptotically optimal, and will likely have extremely good performance throughout the range of sample sizes. This is an area we are exploring, especially as regards 2-stage designs.

Finally, to summarize our findings, we have developed various designs for a clinical trial model with Bernoulli observations, exponentially delayed response times, and Poisson patient arrival rates. Regarding optimal designs, we examined two approaches to tackling the problem and discussed the computational difficulties associated with each, showing that one of them is amenable to implementation. We found that under fairly broad circumstances, in a delayed setting, using the optimal delayed response design gives results nearly as good as those for the immediate response setting when the optimal design is used. We also found that the most commonly proposed ad hoc rule for such problems, the DRPW rule, performed significantly less well than the optimal delayed design, while a new ad hoc strategy, the DAH rule, performed extremely well.

Acknowledgements

This work was partially supported by National Science Foundation grant DMS–0072910. Parallel computing facilities were provided by the University of Michigan's Center for Advanced Computing.

References

[1] Bai, Z., Hu, F. and Rosenberger, W., 2002. Asymptotic properties of adaptive designs for clinical trials with delayed response. Ann. Statist. 30, 122–139.

[2] Bandyopadhyay, U. and Biwas, A., 1996. Delayed response in randomized play-the-winner rule: a decision theoretic outlook., Calcutta Statist. Assoc. Bul. 46, 69–88.

[3] Biswas, A., 2003. Generalized delayed response in randomized play-the-winner rule., Comm. Stat.–Sim. and Comp. 32, 259–274.

[4] Colton, T., 1963. A model for selecting one of two medical treatments. J. Amer. Statist. Assoc. 58, 388–400.

[5] Douke, H., 1994. On sequential design based on Markov chains for selecting one of two treatments in clinical trials with delayed observations. J. Japanese Soc. Comput. Statist. 7, 89–103.

[6] Eick, S., 1988. The two-armed bandit with delayed responses. Ann. Statist. 16, 254–264.

[7] Hardwick, J., Oehmke, R. and Stout, Q.F., 1999. A program for sequential allocation of three Bernoulli populations. Comput. Stat. and Data Analysis 31, 397-416.

[8] Hardwick, J., Oehmke, R. and Stout, Q.F., 2001. Optimal adaptive designs for delayed response models: exponential case. In: A. Atkinson, P. Hackl, W. Mller (Eds.) MODA6: Advances in Model-Oriented Design and Analysis, Physica Verlag. Heidelberg, 127-134.

[9] Hardwick, J. and Stout, Q.F., 1998. Flexible algorithms for creating and analyzing adaptive sampling procedures. New Developments and Applications in Experimental Design, IMS Lec. Notes–Mono. Series 34, 91–105.

[10] Langenberg, P. and Srinivasan, R., 1982. On the Colton model for clinical trials with delayed observations — dichotomous responses. Biometrical J. 24, 287–296.

[11] Oehmke, R., Hardwick, J. and Stout, Q.F., 2001. Scalable algorithms for adaptive statistical designs. Scientific Prog. 8, 183-193.

[12] Rosenberger, W., 1999. Randomized play-the-winner clinical trials: review and recommendations. Cont. Clin. Trials 20, 328–342.

[13] Tantiyaswasdikul, C., 1992. Isotonic Regression under Sequential Designs, Ph.d. thesis, U. Michigan.

[14] Wang, X., 2000. A Bandit process with delayed responses. Stat. and Prob. Letters 48, 303–307.

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

A Appendix: Equations for Approach I

For the patient arrival based approach, at each state $\sigma = \langle s_1, f_1, u_1; s_2, f_2, u_2 \rangle$ there only two options, namely allocating the new patient to arm 1 or arm 2. If we allocate the patient to arm i, we can think of this as initially increasing u_i by one. By the time the next patient arrives, some number of past patients

allocated to arms 1 and 2 will have had their responses. Our goal is to pick the better of the two options. Thus we are led to the following equation:

$$V(\sigma) = \max \left\{ \begin{array}{l} \displaystyle\sum_{j_1=0}^{\widehat{u_1}} \sum_{j_2=0}^{u_2} \sum_{s_1'=0}^{j_1} \sum_{s_2'=0}^{j_2} t(j_1, \widehat{u_1}, j_2, u_2)\, q_1(s_1' \mid j_1)\, q_2(s_2' \mid j_2) \;\cdot \\ \qquad\qquad V(\langle s_1+s_1', f_1+j_1-s_1', \widehat{u_1}-j_1; s_2+s_2', f_2+j_2-s_2', u_2-j_2\rangle)\;, \\ \displaystyle\sum_{j_1=0}^{u_1} \sum_{j_2=0}^{\widehat{u_2}} \sum_{s_1'=0}^{j_1} \sum_{s_2'=0}^{j_2} t(j_1, u_1, j_2, \widehat{u_2})\, q_1(s_1' \mid j_1)\, q_2(s_2' \mid j_2) \;\cdot \\ \qquad\qquad V(\langle s_1+s_1', f_1+j_1-s_1', u_1-j_1; s_2+s_2', f_2+j_2-s_2', \widehat{u_2}-j_2\rangle) \end{array} \right\}$$

The value $t(j_1, u_1, j_2, u_2)$ represents the probability that exactly j_i of the u_i outstanding responses occur from arm i prior to the next patient arrival, $j_i = 0, \ldots, u_i$, and $q_i(s \mid j)$ is the probability that exactly s of the j responses on arm i are successes, $s = 0, \ldots, j$, $i = 1, 2$, given the priors and observations to date. Thus,

$$t(j_1, u_1, j_2, u_2) = P\left(\max\{Z_{\lambda_1(j_1)}, Z_{\lambda_2(j_2)}\} < Z_{\lambda_s} < \min\{Z_{\lambda_1(j_1+1)}, Z_{\lambda_1(j_2+1)}\}\right)$$

where $Z_{\lambda_s} \sim \exp(1/\lambda_s)$, $Z_{\lambda_i} \sim \exp(1/\lambda_i)$ and $Z_{(j)}$ represents the j^{th} order statistic.

While the t, q_1, and q_2 values can be computed once and stored, there is still the difficulty that determining $V(\sigma)$ depends on $O(n^4)$ other states. Since there are $\Theta(n^6)$ states, a straightforward approach to solving all of the recurrences will grow as $\Theta(n^{10})$. While it might be possible to reduce this some, it will likely remain infeasible for useful values of n.

B Appendix: Computations for Approach II

Due to its high dimensional nature, programming the recurrence in Approach II can be somewhat challenging. The state space is 6-dimensional, so the most straightforward implementation would use an array of size n^6. By using the well-known techniques (see Hardwick and Stout, 1998), of doing calculations level by level and overwriting old values this array space can be reduced to n^5, and by utilizing the constraint that $s_1 + f_1 + u_1 + s_2 + f_2 + u_2 \leq n$ and mapping to a 1-dimensional array, this can be further reduced to $\binom{n+5}{5} \approx n^5/5!$. For sample sizes of size 100, however, this is still near the limit of standard computers. Here we will merely sketch some of the programming considerations, especially those aspects that differentiate the D2AB from fully sequential bandits.

As noted above, to reduce the memory needed one should do calculations level by level, starting at the terminal states and proceeding towards the beginning, ending when $V(0,0,0;0,0,0)$ has been determined. While the state space for the delayed 2-armed bandit is in 1-1 correspondence with that for the fully sequential 3-armed bandit, the recurrence for the D2AB is slightly more complex, which complicates the level by level approach. The D2AB recurrence is of the form

$$V(\sigma) = f\left(V(\sigma+\widehat{s_1}-\widehat{u_1}), V(\sigma+\widehat{f_1}-\widehat{u_1}), V(\sigma+\widehat{u_1}), V(\sigma+\widehat{s_2}-\widehat{u_2}), V(\sigma+\widehat{f_2}-\widehat{u_2}), V(\sigma+\widehat{u_2})\right) \quad (1)$$

while in the standard 3AB the recurrence is of the form

$$V'(\sigma) = f'\left(V'(\sigma+\widehat{s_1}), V'(\sigma+\widehat{f_1}), V'(\sigma+\widehat{s_2}), V'(\sigma+\widehat{f_2}), V'(\sigma+\widehat{s_3}), V'(\sigma+\widehat{f_3})\right).$$

The most critical difference is that in the D2AB, the u_1 and u_2 indices are both incremented and decremented. For the standard 3AB, the level is merely the sum of the components, and it is easy to see that each state depends only on the states of level one greater. While this definition of level will not work for the D2AB, by defining the level of a state to be $2(s_1 + f_1 + s_2 + f_2) + u_1 + u_2$, one sees that in equation 1, the value of a state at level ℓ depends only on states at level $\ell + 1$. Thus the level by level approach can be used, from level $2n$ down to level 0.

Unfortunately, this definition of level slightly complicates the programming. For a level $\ell > n$, one can have nonnegative values of s_1, f_1, u_1, s_2, f_2, and u_2 which map to level ℓ, but do not correspond to a valid state because they have a sum that exceeds n. Therefore one must check both lower bounds, as well as upper bounds, on the loops. This problem also complicates the indexing when the state space is mapped into a 1-dimensional array. Overall, the details are considerably messier than for fully sequential bandits, but the same basic approaches can be used. Figure 3 provides a sketch of a serial algorithm for the D2AB.

Note that one can provide exact analyses of arbitrary designs, whether created for delayed responses or not, by merely replacing the "$\max\{V(\sigma+\widehat{u_1}), V(\sigma+\widehat{u_2})\}$" component in the recursive calculation of $V(\sigma)$ with the choice that the design would make. Further, one can optimize or evaluate with respect to arbitrary objectives, not just total successes, by replacing "$V(\sigma)$=number of successes in σ" with the appropriate value of the new objective function at σ.

The parallel program for the D2AB starts with the serial program and then divides the work among the processors. The m loop cannot be parallelized, but the loops within it can be. One must also add communication among the processors so that each processor has the V values it needs from the previous $m + 1$ iteration to determine its V values for the current m iteration. See Hardwick et al. (1999) and for a more detailed discussion of the parallelization process. Also, see Oehmke et al. (2001) for more detailed timing analyses and optimizations to improve serial and parallel performance. Problems as large as $n = 200$ have been solved using modest parallel computers, ones that are widely available.

$\{\widehat{s_i}, \widehat{f_i}$: one success, failure on arm i $\}$
$\{s_i, f_i, u_i$: number of successes, failures, unobserved on arm i $\}$
$\{\pi_i$: posterior probability of success on arm i $\}$
$\{n$: sample size $\}$
$\{m$: level $\}$
$\{V(\sigma)$: position in 1-dimensional array of value of state σ $\}$
$\{V_i$: expected value if observation occurs on arm i $\}$

forall states σ with level(σ)=$2n$ {i.e. for all terminal states}
    $V(\sigma)$=number of successes in σ

for $m = 2n-1$ downto 0 {compute for all states of level m}
    for $s_2 = 0$ to $m/2$
        for $f_2 = 0$ to $(m/2) - s_2$
            for $s_1 = 0$ to $(m/2) - s_2 - f_2$
                for $f_1 = \max(0, m-n)$ to $(m/2) - s_2 - f_2 - s_1$
                    for $u_2 = 0$ to $m - 2(s_2 - f_2 - s_1 - f_1)$
                        $u_1 = m - 2(s_2 - f_2 - s_1 - f_1) - u_2$
                        $\sigma = \langle s_1, f_1, u_1; s_2, f_2, u_2 \rangle$
                        $V_1 = \pi_1(s_1, f_1)V(\sigma + \widehat{s_1} - \widehat{u_1}) + (1 - \pi_1(s_1, f_1))V(\sigma + \widehat{f_1} - \widehat{u_1})$
                        $V_2 = \pi_2(s_2, f_2)V(\sigma + \widehat{s_2} - \widehat{u_2}) + (1 - \pi_2(s_2, f_2))V(\sigma + \widehat{f_2} - \widehat{u_2})$
                        if $s_1 + f_1 + u_1 + s_1 + f_2 + u_2 < n$ **then** {more subjects possible}

$$V(\sigma) = \frac{\lambda_s \max\{V(\sigma + \widehat{u_1}), V(\sigma + \widehat{u_2})\} + u_1\lambda_1 V_1 + u_2\lambda_2 V_2}{\lambda_s + u_1\lambda_1 + u_2\lambda_2}$$

                        else

$$V(\sigma) = \frac{u_1\lambda_1 V_1 + u_2\lambda_2 V_2}{u_1\lambda_1 + u_2\lambda_2}$$

Figure 3: Outline of serial program for delayed 2-armed bandit