Adaptive Designs
IMS Lecture Notes - Monograph Series (1995) Volume 25

A MODIFIED BANDIT AS AN APPROACH TO ETHICAL ALLOCATION IN CLINICAL TRIALS

BY JANIS P. HARDWICK*

The University of Michigan

Abstract

A sequential allocation rule based on an optimal strategy for a two-armed bandit problem is proposed for use in the problem of identifying the better of two treatment alternatives in clinical trials. The purpose of the rule is to ensure more ethical allocation of patients while retaining a given probability of correctly selecting the better treatment at the close of the trial. The behavior of the bandit rule is compared with two other commonly proposed allocation rules: play-the-winner and vector-at-a-time. It is found that, in general, the bandit rule performs as well as, and usually remarkably better than, both of the other allocation rules. All comparisons are based on exact computations using forward induction algorithms carried out on desktop workstations.

1. Introduction. Medical research is often complicated by the inherent dependence of biological experimentation on living subjects. Moral conflicts are invariably generated by the obligation of a researcher

Received January 1993; revised April 1993.

*Research supported in part by the National Science Foundation under grants DMS-8914328 and DMS-9157715.

AMS 1991 subject classification. Primary 62L05, 62L10, 62L13, 62F15.

Key words and phrases. Adaptive allocation, Gittins index, sequential allocation, two-armed bandit.

to balance the well-being of subjects who participate in clinical trials with that of future patients who stand to benefit from new advances in medical treatment. This longstanding dilemma has received considerable attention in both the statistical and medical literatures, where discussions pit utilitarian ideals against the individualistic ethic [see Anscombe (1963); Bartlett, Roloff, Cornell, Andrews, Dillon and Zwischenberger (1985); Byar, Simon, Friedewald, Schlesselman, DeMets, Ellenberg, Gail and Ware (1976); Weinstein (1974)]. As a consequence, the problem of determining the better of the treatment alternatives must encompass the consideration of *ethical* factors as well as the costs of incorrect population selection and sample size.

Classical randomized clinical trials in which patients are allocated in approximately equal proportions to different treatment regimens represent an extreme of the utilitarian goal. The individualistic goal, on the other hand, is exemplified in trials employing *myopic* allocation, in which patients are assigned to the treatment that has the highest current expected probability of success.

Allocation rules in which patients are *adaptively* assigned to competing therapies offer a compromise between the two extremes. In the early stages of sampling, adaptive schemes emulate proportional rules, and later, if the data suggest that one of the treatments is superior, the rules adjust themselves in the direction of myopic allocation.

Frequently, a main objective of a clinical trial will be to determine whether a new treatment is better than a current favorite or a competing new treatment. In particular, the ability to make decisions based on classical frequentist arguments seems to be part of accepted clinical trial methodology. Adaptive sampling strategies are not designed to collect information that leads naturally to these types of inferences. A reasonable question, then, is whether investigators can attain their research goals while simultaneously promoting the welfare of the patients within the trial. In this paper, we show that, when applied in appropriate research settings, an adaptive clinical trial design can not only provide information commensurate with that in a balanced design, but also reduce costs to patients associated with the trial.

2. The problem. Consider data collected within the framework of a controlled trial into which patients are entered sequentially and assigned to one of two treatments. An upper bound is placed on the number of subjects in the trial, N, but, because one may stop before N patients have been entered, the actual sample size for a trial is a

random variable, n. With regard to the patient responses, we assume the following:

- The outcome measures are z_i, $i = 1, \ldots,$ where $z_i =$ and x_{ji}, $j = 1, 2$, are independent Bernoulli random variables with success parameters P_j; $(P_1, P_2) \in \Omega = (0,1) \times (0,1)$.
- The outcome for patient i is known before assignment is made for patient $i + 1$.
- The success parameters, P_1 and P_2, remain constant throughout the trial.

2.1. Error probabilities. As stated, the research goal for the trial is to determine the better of the two treatments, so a primary concern in the design of the trial is that there be a high probability of arriving at a correct terminal decision. Even if there were *no* other concerns or costs associated with the trial, it would be impossible to choose a single design that would optimize the *probability of correct selection*, $P(CS)$, for all $(P_1, P_2) \in \Omega$. Thus, to make the problem tractable, we restrict the class of designs that we consider to include only those that satisfy the following constraint: We must have

(1)
$$P(CS) \geq P^* \quad \text{whenever} \quad |P_2 - P_1| = \Delta \leq \Delta^*.$$

The constants P^* and Δ^* are chosen in advance to satisfy the needs of the medical researchers and are known in the selection literature as the (P^*, Δ^*) requirement [see, for example, Bechhofer, Kiefer and Sobel (1968)]. Other parameters of the selection design are chosen to satisfy the (P^*, Δ^*) requirement along with all other design criteria.

2.2. Other criteria. We consider designs with three components here - an allocation rule, a stopping rule and a terminal decision rule. Our main interest, however, is to isolate allocation rules so that comparisons may be drawn among them. To this end we fix the form of the stopping and terminal decision rules.

A terminal decision rule for the two-population selection problem has three possible outcomes: select one of the two treatments as better or declare no meaningful difference between treatments. Here, we focus primarily on the *two* outcome decision rule that specifies a single 'better' treatment at the trial's termination. However, since selection

procedures have often been criticized for disallowing the reasonable conclusion that there exist times when one should not differentiate based on the available data, we include results for the three decision model as well. We believe the general behavior of $P(CS \mid P_1, P_2)$ for this model provides some extra insights regarding the allocation rules of interest.

We base our stopping and terminal decision rules on the criteria used in Kiefer and Weiss (1971, 1974), as theirs were among the rare studies that allowed the 3-way rule and it is useful to have an external basis for comparing some of our calculations.

For Bernoulli outcome variables, it is common to use a stopping rule that depends on the number of accumulated successes, S_{ik}, and/or failures, F_{ik}, for the two treatments, $i = 1, 2$ at Stage $k \leq N$. Following Kiefer and Weiss (1971, 1974), we consider stopping rules based only on the number of observed successes. The main component of our stopping rule calls for termination whenever the difference between the observed number of successes equals a predetermined constant, r, that is, stop at Stage k if
(2)

$$\mid S_{1k} - S_{2k} \mid = r, \quad r \leq k \leq N.$$

A secondary feature of the rule is *curtailment.* To curtail a trial means that we stop it as soon as there is sufficient information to guarantee that the treatment selected as *better* could not be altered by the remaining number of observations - regardless of their outcomes. Thus, as soon as we reach a curtailment point or (2) occurs, the trial is stopped and the treatment with the greater number of successes is declared the winner. If $S_{1N} = S_{2N}$, then a fair coin is tossed to determine the better treatment.

To incorporate a decision of *no difference* for the three way decision rule, we again consider curtailed rules. That is, we stop and select a better treatment if we encounter the event (2); but if, at some stage k, $N - r \leq k \leq N$, the difference $\mid S_{1k} - S_{2k} \mid$ cannot possibly reach r in the remaining trials, we stop and declare no meaningful difference between the treatments. For both stopping and decision structures, the parameters r and N are chosen to satisfy the (P^*, Δ^*) requirement.

2.3. Ethical criteria. Clinical trials involve multiple costs, so a good trial will be designed to diminish whatever combination of costs seem to be most critical. In some instances, the primary costs may

be financial. In others, time may be the major concern. Here, we are interested in *ethical* costs. There are a variety of ways to measure the ethical costs associated with a clinical trial, and we do not argue here for the usage of any particular cost function over another. That is a different discussion entirely, and we refer readers to Bather (1985), Anscombe (1963), Berry and Eick (1987), Chernoff and Petkau (1985), and Hardwick (1989), among others.

In this study, the goal is to achieve some sort of fair comparison among allocation rules. In general, such comparisons entail the study of the number of patients who do poorly during the study, the number of patients required to carry out the study, and the number of 'extra' patients who do poorly in the future as a result of an 'incorrect' terminal decision. In the present set-up, the latter factor is addressed mainly by the (P^*, Δ^*) requirement. As for the former two criteria, we consider the following measures:

- **Expected Successes Lost**$[E(SL)]$: *Consider a design A and let $N_H \geq N$ be an extended fixed horizon size for the problem. For this measure, we carry out the trial as usual, but when a terminal decision is made at time $n \leq N$, we allocate the patients from Stage $n+1$ to Stage N_H to the treatment declared as better at Stage n. If $E(S \mid A, N_H)$ is the total expected number of successful patient outcomes in the trial of N_H observations (using design A), then the* expected successes lost *is given by*

$$E(SL \mid A, N_H) = \max(P_1, P_2) \times N_H - E(S \mid A, N_H).$$

- **Expected Failures**$[E(F)]$: *The number of failures that one expects to observe before arriving at a terminal decision.*

- **Expected Inferior Treatment**$[E(I)]$: *The expected number of patients allocated to the inferior therapy during the trial.*

- **Expected Sample Size**$[E(n)]$: *The expected study length.*

Given two designs, A and B, with given stopping and terminal decision rules, we define design A to be *superior* to design B with respect to criterion $\mathcal{C}$ in a region $\mathcal{R}$ of the parameter space if

- A and B both satisfy the (P^*, Δ^*) requirement, and

Note that the parameters that allow different procedures to satisfy the (P^*, Δ^*) requirement may differ among procedures.

3. Play-the-winner and vector-at-a-time allocation rules. Two allocation rules that have received significant attention in the literature are Play-the-winner (PW) and Vector-at-a-time (VT). As originally described by Robbins (1952), the PW rule is an adaptive procedure whose strategy indicates that the investigator follow a successful trial outcome by another trial from the same population, and follows an unsuccessful outcome by a trial from the opposite population. There are many variations of the PW rule [see, for example, Zelen (1969), Hoel, Sobel, and Weiss (1975), and Wei and Durham (1978)], but we concentrate on the above version because of its intuitive appeal and tractability.

Vector-at-a-time (VT) sampling is the usual balanced setup for matched pair studies in which trials are carried out two at a time with one observation being made from each of the two treatments at each stage of the trial. Equal allocation strategies such as VT are used often in practice because they have excellent decision making properties, circumvent objections to unbalanced allocation, and are well understood analytically.

For two-population dichotomous outcome problems, numerous studies have been carried out under the constraints (1), and studies of the Play-the-Winner and Vector-at-a-Time rules have dominated this literature. In their book, Buringer, Martin, and Schriever (1980) compile and discuss a large number of these studies noting repeatedly that there are certain regions of the parameter space in which one or the other of the rules PW or VT consistently outperforms the other. They also point out that under virtually no circumstances will either rule behave better *uniformly* for $(P_1, P_2) \geq \Delta^*$, let alone for all $(P_1, P_2) \in \Omega$. We are interested in finding an allocation rule that mimics the behavior of each of these two rules in the region of the parameter space in which they perform the best.

4. Bandit allocation rules. Much of the early work in the area of adaptive allocation was done in the context of what are termed as *Bernoulli two-armed bandit* (TAB) problems. In such problems, the investigator observes, at each stage, a Bernoulli random variable from either Population 1 (with unknown success rate P_1) or Population 2 (with unknown success rate P_2), with the goal of maximizing the sum

of an infinite sequence of observations that are discounted according to some sequence $\mathcal{B} = (\beta_1, \beta_2, \beta_3, \ldots)$, $0 \leq \beta \leq 1$.

Originating with Thompson (1933) and finding early application in research on adaptive learning, TAB problems pose a significant challenge to investigators who wish to understand the tradeoffs inherent in problems where information may lead to reward. Since the 1930's, numerous researchers have studied properties of the TAB. Berry and Fristedt (1985) provide a summary of key results in their book *Bandit Problems: Sequential Allocation of Experiments.* Their book also contains an annotated bibliography offering a broad and excellent overview of bandit related literature.

Interestingly enough, even in its first incarnation, the TAB problem was posed as a problem of allocating patients within a clinical trial [see Thompson (1933)]. Thompson considered a finite horizon trial with N patients in which each response carried the same weight. This is equivalent to using a *uniform* discount sequence of length N : $\mathcal{B} = (\underbrace{1, 1, \ldots, 1}_{N}, 0, 0, \ldots)$. By placing independent uniform priors on P_1 and P_2, he then studied the TAB problem within a Bayesian framework.

Bayesian formulations of bandit problems are now the norm; and, while the choice of prior distribution may vary according to the investigator, we follow Thompson and use independent uniform priors. These priors are useful for initial comparisons among allocation rules because 1) they contain little initial bias, 2) the prior information is rapidly overcome by the information from incoming data, and 3) the parameters of the beta posterior distributions concisely summarize the relevant information from the study to date. It is certainly the case, however, that there may be situations in which the prior distributions should reflect a greater amount of information or even a bias in favor of one of the therapies. In such cases, a beta prior with carefully chosen parameters can often provide a reasonable representation of an investigator's prior knowledge without increasing the complexity of the problem [see Hardwick (1986a)].

We refer to the bandit problem with uniform discount sequence and independent uniform priors as the *uniform* TAB (UTAB). While more or less inaccessible in 1933, optimal solutions for the UTAB do exist, and with currently available computers, can be obtained via dynamic programming.

In the UTAB problem, an optimal strategy will, in the early stages, emphasize the gathering of information, possibly at the expense of the

gathering of immediate reward. The effect will be that the initial patients will be treated like patients in a VT design, in which it is assumed until the end of the study that the treatments offer the same prognosis. Toward the end of the study, when a decision is imminent and the need for information is diminished, the emphasis on immediate reward is increased until, at the last stage, the myopic rule is used. During the middle stages, there is a gradual shifting of emphasis from the accumulation of information to that of reward. An intriguing feature of the optimal strategy is that it neither treats trial patients uniformly nor does it optimize the treatment of each patient individually. Thus, even a rule that optimizes the treatment of the trial patients *on average* retains some conflicting ethical attributes. See Berry (1978) for applications of the UTAB in clinical trials designs.

In an attempt to duplicate some of the positive ethical features of the UTAB while alleviating some of the obstacles it poses, we turn to an allocation rule based on a bandit problem that utilizes a different discount sequence. The rule we examine, the *modified* bandit (MB), is an ad hoc rule based on optimal strategies for a TAB with *geometric* discount (GTAB). In the GTAB with discount sequence, $\mathcal{B} = \{1, \beta, \beta^2, \beta^3, \ldots\}, 0 < \beta < 1$, the goal is to maximize the expectation of the discounted sum of observations, $\Sigma_1^\infty \beta^{i-1} z_i$, over an *infinite* horizon, rather than the finite sum, $\Sigma_1^n z_i$ encountered in the UTAB.

The optimal strategy for this bandit problem has a special appeal for clinical trial applications since it can be viewed as offering an ethically equitable mechanism for balancing outcomes of present and future patients: *the relative weight of the result of each patient is constant when compared to the sum of the weights of all future patients.* In the GTAB, one assumes that there will always be more patients, so the need for information is never completely absent as it is in the last stage of a finite horizon problem. However, as more and more patients are treated, information accumulates, and thus, the need to sacrifice immediate reward to gain information is diminished. Note that, while here again, the allocation rule does not attempt to treat each patient optimally, it does seek to treat them all the same, while simultaneously seeking to maximize the cumulative well being of the patients.

As mentioned, bandit allocation strategies are not designed to be used for statistical inference, so there is no inherent (inferential) decision making structure that arises from using such rules. To account for the fact that the geometric bandit has an infinite horizon, we overlay it with the stopping rules relevant to the problem. Further, to accom-

modate the need to make a decision, we append the design with the terminal decision corresponding to the stopping rule selected.

Thus, the rule that we study here is a finite horizon rule that behaves as an optimal allocation rule for a geometric TAB until either the stopping rule or the truncation rule is invoked. Because the eventual forced stopping is not incorporated into the decision process as initial information (as it would be in a non-modified finite horizon TAB problem) the MB continues to seek information even during the final stage of the trial, just as physicians are never finished learning from their patients. Thus, while the modification of the bandit rule is due to the practical constraint of eventually having to draw a conclusion regarding the two treatments, it is actually a rather realistic tactic.

5. Computational methods and results. Ethical attributes aside, an experimental design must be straightforward to carry out if it is to be useful. PW and VT are trivial procedures to define and follow, and we know optimal solutions for the UTAB can be obtained with a great deal of computation. The GTAB problem, however, is an *infinite horizon* problem, and such problems cannot be solved via the dynamic programming methods used for finite horizon TAB problems. In fact, exact solutions to GTAB's would be all but impossible were it not for a remarkable result due to Gittins and Jones (1974).

According to the Gittins and Jones theorem, in a bandit problem with geometric discount and independent arms, there exists for each arm separately an index with the property that, at any given stage, n, it is optimal to allocate the next patient to the treatment with the higher index. The index for an arm, the Gittins Index, may be defined in terms of a *one-armed* bandit [see Berry and Fristedt (1985)].

In the one-armed bandit problem, there is an arm with unknown success probability, P_0, competing against an arm with known success probability, P_k. Let the prior distribution of P_0 be fixed. If P_k is sufficiently small, it will be optimal to use the unknown arm at the next stage. If P_k is sufficiently large, it will be optimal to use the known arm at the next stage and for all successive stages. The Gittins Index, associated with the arm with unknown success probability, is defined to be the value of P_k such that one would be indifferent if one had to choose between the known and unknown arms at the next stage. The index is a function of the posterior distribution and the discount factor β.

While the existence of the Gittins Index removes many computa-

tional difficulties associated with the GTAB, actually computing the index for any but the most trivial of bandit problems remains problematic. Furthermore, the number of operations involved in the computation of the index increases drastically as β approaches one. Since we have found that the most effective values of β for our application are those that approach one rapidly as the horizon size increases, we use approximations for the index that are far easier to compute.

The approximations that we use, due to Berry and Fristedt (1985), are actually lower bounds for the Gittins Index. Since we are interested only in the relative values of the indices, the fact that these approximations are everywhere lower bounds is an asset. Furthermore, while the lower bounds are close to the true index [see, for example, Table 5.2 of Berry and Fristedt (1985)], it is not clear that using the index itself would improve performance for our application. From the general expression for the lower bound given an arbitrary prior distribution [example 5.4.6 of Berry and Fristedt (1985)], Hardwick (1986b) derives the lower bound for the special case in which the prior distribution is beta. This bound with geometric discount, $\{\beta^i\}$, and beta prior with parameters (A,B) is given by $\Lambda_* = \sup\{\Lambda_r : r = 1, 2, \ldots\}$, where

$$\Lambda_r = \frac{\Gamma(A+1)/\Gamma(A+B+1) - B\Sigma_1^r \beta^i \Gamma(A+i)/\Gamma(A+B+i+1)}{\Gamma(A)/\Gamma(A+B) - B\Sigma_1^r \beta^i \Gamma(A+i-1)/\Gamma(A+B+i)}.$$

Since $\{\Lambda_r : r = 1, 2, \ldots\}$ is unimodal, we may let $r^* = \min\{r : \Lambda_{r+1} \leq \Lambda_r\}$, and then set $\Lambda_* = \Lambda_{r^*}$.

Aside from the extra computational work associated with *defining* certain adaptive procedures, there is also the problem of determining the characteristics of such procedures. With a rule as simple as PW, one can often obtain design characteristics analytically, as was done by Kiefer and Weiss (1974). More frequently, however, investigators find it necessary to simulate or otherwise approximate the performance of adaptive procedures in order to examine properties of the rules for an interesting range of parameter configurations. While such studies have been critical to our developing understanding of adaptive designs, the advent of better computing algorithms and faster, more powerful computers allows us to compute exactly many of the results that in the past, we could only approximate. In particular, the results presented here have been obtained using forward induction or *path counting* algorithms as described by Hardwick (1991) and the paper by Hardwick and Stout in this volume, and were carried out on a desktop workstation.

As mentioned in Section 3, there exist no rules that perform uniformly best (over all Ω) with respect to $P(CS)$. Furthermore, even after restricting the class of rules to those satisfying the (P^*, Δ^*) requirement, it is difficult to locate families of rules that are uniformly best with respect to a second criterion such as $E(SL), E(F), E(I)$ or $E(N)$. Since our goal here, however, has been to study the dynamics of the rules with respect to all of the specified criteria, we have sought design parameters (r, n) and β that perform well in general. Naturally one could emphasize a single criterion, as did Kiefer and Weiss (1971, 1974) with $E(I)$, tuning the design parameters for that criterion alone.

Regarding the general behavior of the VT, PW, and MB rules, our comments are based on data from numerous (P^*, Δ^*) combinations. The tables and figures included here are drawn from data sets chosen to exemplify the trends we observed.

5.1. The two-decision problem. The first data set was computed for the *two decision* problem with $(P^* = 0.9, \Delta^* = 0.1)$, and Figures 1 through 9 provide graphical representations of the results. The parameter configurations for the designs displayed in the figures is $(P^* = 0.9, \Delta^* = 0.1)$, with $\beta = 1 - 10^{-11}$ and

$$\begin{array}{lll} \textbf{VT}: & r = 13 & \mathbf{N = 164} \\ \textbf{PW}: & r = 20 & \mathbf{N = 165} \\ \textbf{MB}: & r = 23 & \mathbf{N = 177} \end{array}$$

Each of the Figures 1 to 4 compares the designs on one measure of ethical quality by indicating the values of the relevant measure at representative points along the line $(P_2 = P_1 + \Delta^*)$. The lines dashed with V's, P's and M's depict the data for the VT, PW and MB rules, respectively. In this example, the horizon N_H for the measure E(SL) is taken to be 250 for all three rules.

First we consider VT, a design whose attributes are largely foreshadowed by our knowledge of its fixed sample size properties. We know, for example, that even though there is a stopping rule in the present design, we should expect symmetry in Figures 1, 2 and 4. Naturally, this rigidness has both disadvantages and advantages. One advantage is simplicity. The characteristics of VT are fairly simple to compute analytically. Furthermore, no computations are required to determine how to allocate the next patient. With respect to performance, however, VT is the least good of the three rules. It offers marginal improve-

Figure 1.
Expected Successes Lost

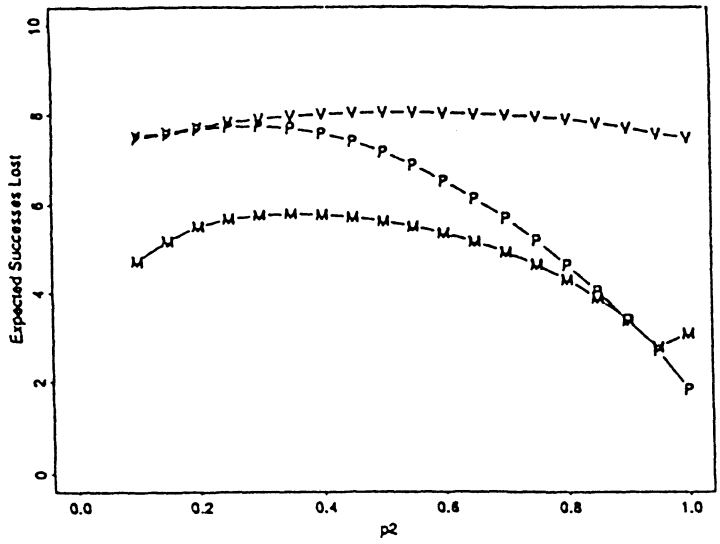


Figure 2.
Expected Inferior Treatment

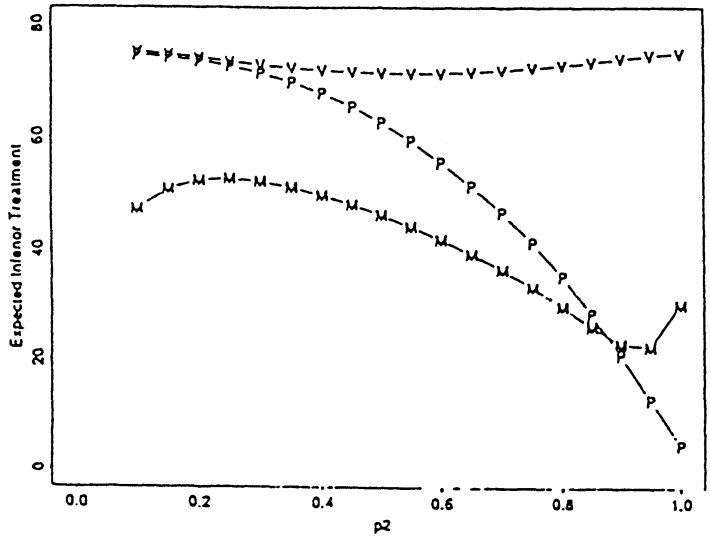


Figure 3.
Expected Failures

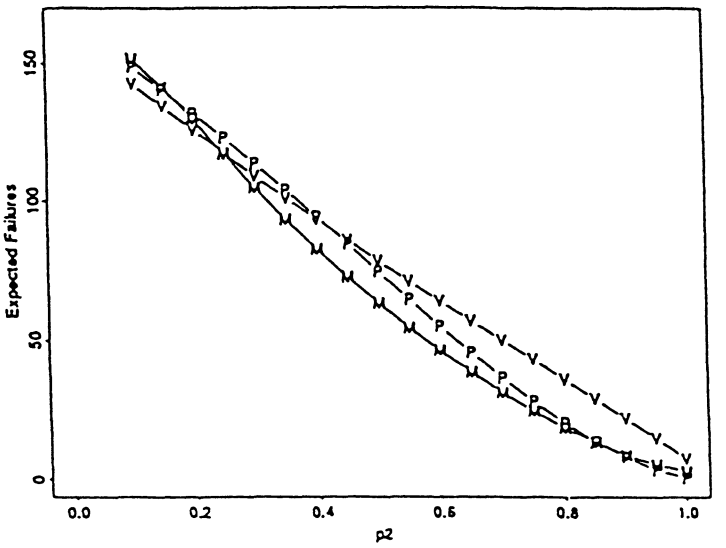
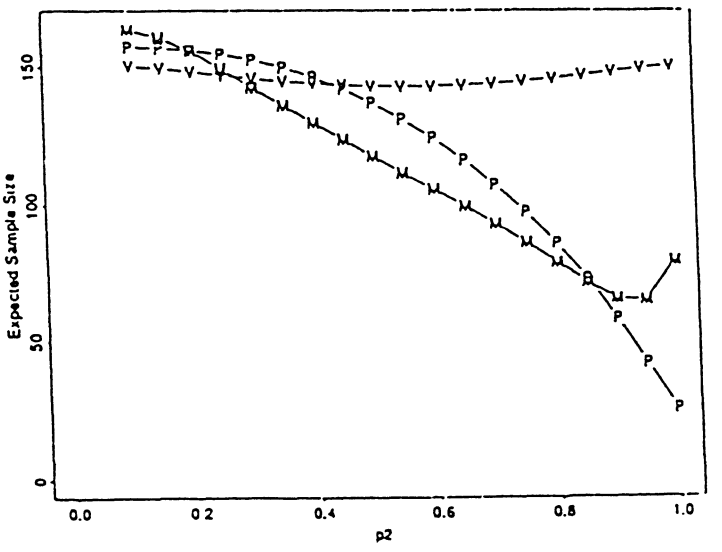


Figure 4.
Expected Sample Size



ment for the variables E(N) and E(F) when $P_2 < .2$, but otherwise it is uniformly worse (and in many cases markedly so).

The PW rule also performs as expected based on our knowledge from other studies [for example, Buringer, Martin, and Schriever (1980)]. It offers the lowest values for all four criteria whenever P_2 is greater than about 0.9. While VT may improve on PW for $P_2 < .2$, the relative difference is very small compared to the improvements made by PW over VT in the upper regions of the parameter space. Still, with PW looking more or less like VT for small parameter values, we find that this rule doesn't really have the overall performance that we desire.

As we had hoped, the MB performs the best of the three rules over the majority of the parameter space displaying truly superior behavior for E(SL) and E(I) in the regions between $0.2 < P_2 < 0.9$. On another positive note, something we have not seen in these data are longer expected study lengths than one typically associates with adaptive designs. The MB rule requires a substantially smaller sample size on average than does VT.

While the results for the case studied here are typical of our more general findings, the precise intervals on $P_2 = P_1 + \Delta^*$ on which the three designs distinguish themselves will vary according to the (P^*, Δ^*) requirement. A couple of other trends that we noticed were that, for a given value of P^*, the relative amounts by which the MB procedure improves on the others increases with both N and Δ^*. See Hardwick (1986a,b) for more detailed comparison of the rules as (P^*, Δ^*) and N are varied.

So far, we have focussed on the behavior of the characteristics of our three procedures along the line $P_2 = P_1 + \Delta^*$. It is also of interest, however, to study the behavior of the criteria over the entire parameter space. While all procedures may satisfy the global (P^*, Δ^*) restriction, the actual values each procedure exhibits pointwise for each variable is likely to be quite different and can be highly informative.

In Figure 5 we show the behavior of the P(CS) for all three rules along the line $P_2 = P_1 + \Delta^*$ for the same situation as was used for Figures 1 through 4. In this case, the three rules look pretty much the same. As an example of the behavior of P(CS) over the entire parameter space, however, we show $P(CS \mid P_1, P_2)$ for a single rule, PW, in Figure 6. This figure indicates that whenever Δ increases beyond 0.2 or 0.3, the P(CS) increases rapidly to one and remains there. The corresponding figures for MB and VT look almost identical and are not shown here.

Sometimes, it is more useful to see pictures of *differences* between

Figure 5.
Probability of Correct Selection

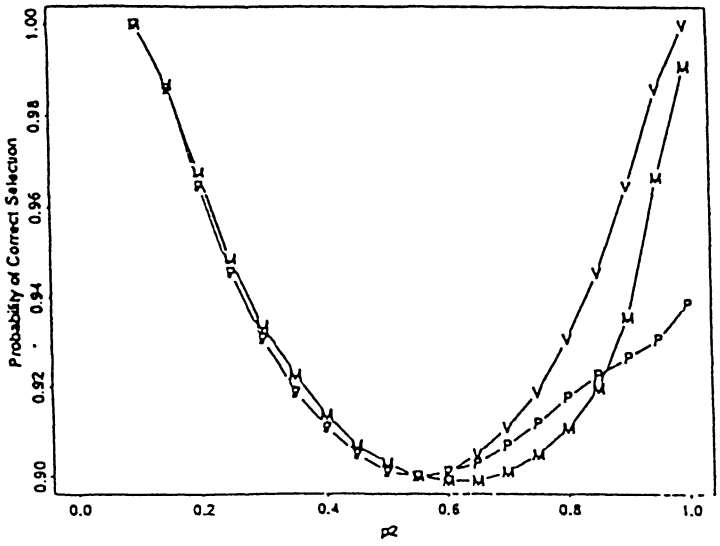


Figure 6.
P(CS) for PW over Ω

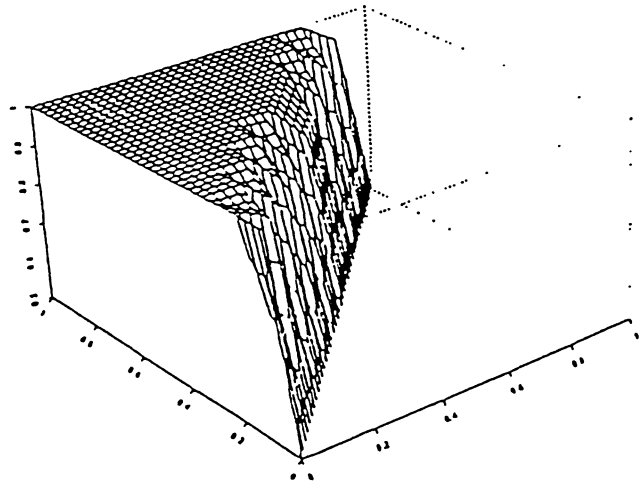


Figure 7.
E(SL]VT) - E(SL]MB)

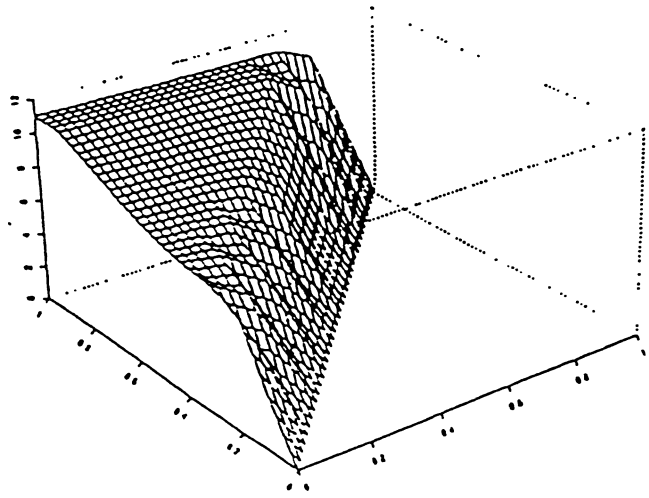


Figure 8.
E(SL]VT) - E(SL]PW)

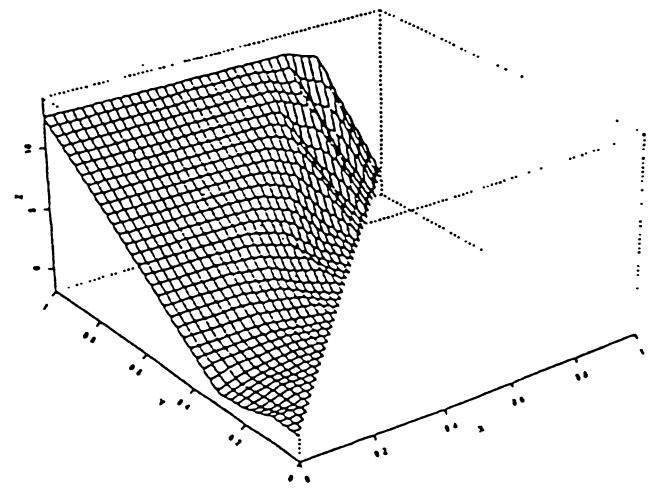
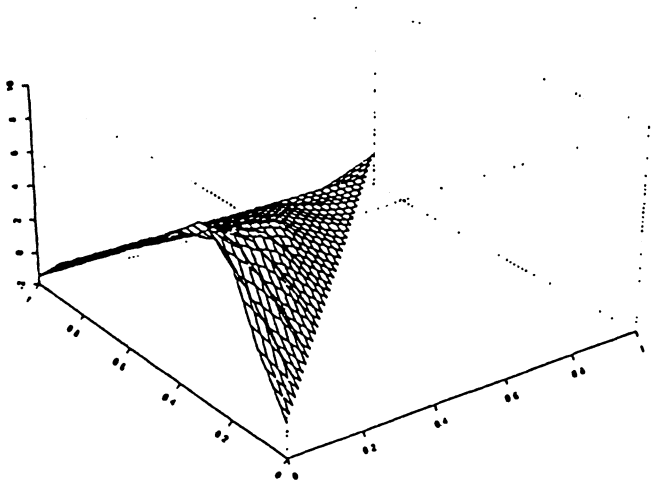


Figure 9.
E(SL]PW) - E(SL]MB)



the values on given ethical criteria for the rules paired up two at a time. Due to space constraints, we only show one set of such pictures here: differences for the variable E(SL). In Figure 7, E(SL|VT) - E(SL|MB) is plotted over the upper triangular portion of Ω. (In all of the three dimensional pictures, the half of the plot not printed is symmetric to the part shown.) From this picture we see that VT is never better than MB for this variable, and that in some regions of Ω, MB is expected to 'lose' as many as 11 fewer patients. Figure 8 offers similar insights for the difference E(SL|VT) - E(SL|PW) except that there appears to be a small region of Ω in which VT improves slightly over PW. In the third difference picture, Figure 9, which shows E(SL]PW) - E(SL]MB), we find that in the average trial with total horizon 250 MB loses significantly fewer patients than PW, but that PW has a minor advantage when one of the parameters is close to one.

5.2. The three-decision problem. For the *three-decision* problem, we work with the data set in which the (P^*, Δ^*) requirement is set at $(0.8, 0.1)$. The data for this example are exhibited in Tables 1 and 2. In Table 1, we have the values for three ethical criteria, E(I), E(F), and E(N), for fixed values along the line $P_2 = P_1 + \Delta^*$; $P_2 = 0.15, 0.25, ..., 0.95$. Table 2 shows the probability of arriving at a terminal decision of *no difference*, P(ND), when $P_1 = P_2$.

We should note that we have restricted the parameter space for the

Table 1.
$P^* = 0.8, \Delta^* = 0.1$

Three-Decision Problem									
P_2	Vector-at-a-Time ($r = 6, N = 180$)			Play-the-Winner ($r = 10, N = 240$)			Modified Bandit ($r = 13, N = 170$) $\beta = 0.999999$		
	E(I)	E(F)	E(N)	E(I)	E(F)	E(N)	E(I)	E(F)	E(N)
0.15	55.3	99.5	110.6	79.0	149.9	167.0	34.7	109.9	125.3
0.25	51.0	81.6	102.0	67.7	114.7	143.9	34.6	78.1	99.5
0.35	47.7	66.8	95.5	57.2	85.5	122.7	31.2	56.5	82.1
0.45	45.8	54.9	91.5	46.7	60.3	101.2	27.0	40.4	68.5
0.55	45.1	45.1	90.2	36.5	39.7	80.2	23.0	28.1	57.3
0.65	45.8	36.6	91.5	27.1	24.0	60.9	19.2	18.7	47.9
0.75	47.7	28.6	95.5	18.9	12.9	44.0	15.4	11.4	39.4
0.85	51.0	20.4	102.0	12.0	5.7	29.9	11.7	5.9	31.6
0.95	55.3	11.1	110.6	6.0	1.5	18.0	8.7	2.2	26.4

three-decision version of the problem. Here we take $\Omega^* = (0.05, 0.95) \times (0.05, 0.95)$. We have done this because the larger r value of the adaptive designs tends to create extreme results as $P_1 \to 0.0$. In this neighborhood, the probability of reaching the decision of 'no difference' is virtually 1.0, and with P(ND) so high, P(CS) is extremely low - forcing the trials to utilize unrealistic numbers of patients.

Note also that if one desired a more typical hypothesis testing design (with both type 1 and type 2 errors), one would not want to use PW or MB rules without making some adjustments, since it is not possible to guarantee a type I error rate, $\alpha = 1 - P(ND \mid \Delta = 0)$, that never exceeds a specified constant, P', over all Ω. Using the present stopping rules, for example, if we consider the point $P_1 = P_2 = 1$, we would have $P(ND) = 0$, because whichever treatment we began with would ultimately be declared the winner. This alone does not concern us too much because we don't think it would be wise to carry out a clinical trial using the current designs when either P_1 or P_2 is thought to be outside of Ω^*. In such cases it often makes more sense to use other functions of the parameters such as ratios or the like.

The bottom line here really is that it seems that the no difference outcome is not an acceptable option for the particular combination of allocation and stopping rules in the designs discussed here. For VT,

Table 2.

$P^* = 0.8, \Delta^* = 0.1;\ \Delta = 0.0$

Probability of Concluding No Difference when $P_1 = P_2$			
P_2	Vector-at-a-Time ($r = 6, N = 180$)	Play-the-Winner ($r = 10, N = 240$)	Modified Bandit ($r = 13, N = 170$) $\beta = 0.999999$
0.05	0.92	0.99	0.99
0.15	0.58	0.76	0.49
0.25	0.40	0.49	0.21
0.35	0.31	0.28	0.10
0.45	0.27	0.14	0.03
0.55	0.27	0.10	0.01
0.65	0.31	0.01	0.00
0.75	0.40	0.00	0.00
0.85	0.58	0.00	0.00
0.95	0.92	0.00	0.00

we would need significantly larger sample sizes to ensure that both the (P^*, Δ^*) requirement *and* an $\alpha \leq P'$ requirement were met. For the adaptive designs, we would need both larger sample sizes and minor adjustments to the rules. Some of these adjustments are outlined below, but there are also a variety of others that one could look into.

First, note in Table 1 that the large values of E(F) for PW in the low range of the parameter space are due to the design's excessive value of E(N) in the region. This latter trait can be attributed to the large value of the stopping parameter r needed to achieve at least $P(CS) \geq P^*$ in the upper parameter ranges. In practice, for small (P_1, P_2), PW sampling is quite similar to VT sampling, switching treatments after each failure. Sampling goes on longer for PW only because the optimal choice of r is higher for this rule than it is for VT. There are a couple of ways to circumvent this problem. To begin, one could utilize a stopping rule such as the present one, but which also incorporates the total number of failures or depends on likelihood ratios. It is known that PW performs better on the average when such stopping rules are used [for example, see Fushimi (1973) and Simon, Weiss and Hoel (1975)]. Since neither type of adjustment to the stopping rule would reduce r, PW would retain its positive features in the upper region of the parameter

space without diminishing P(CS). In the lower region of the space, however, the trial would be stopped once a certain number of failures had been observed. This would guarantee a stopping time more similar to that of VT. For the same reasons, we would expect improvement for the MB, which has problems similar to, but not as extreme as those of PW. Less improvement would be expected for the VT designs since there is symmetry between the numbers of successes and failures for the two treatments.

Another possible approach would be to try a two stage process in which r could depend on current estimates of (P_1, P_2). In this case, one might begin with the value of r defined in the present design. Then, part way through the trial, one could estimate the success rates, and if they both appeared to be below, say 0.5, one could determine a new values of r (and N) that would guarantee only that $P(CS \mid P_1 < 0.5\,, P_2 < 0.5) \geq P^*$.

We have not studied any of the above rules, and recognize that a number of adjustments would need to be made to ensure that the design's overall performance was not diminished. We do believe, however, that it is of interest to see if there's anything simple one can do to the adaptive rules to guarantee their good performance for the three-decision selection problem before abandoning them permanently.

6. Discussion. Our intent has been to show that one can use adaptive allocation schemes to improve the treatment of patients during a clinical trial without interfering with the ability to make correct terminal decisions. Our approach has been to consider a binomial selection model in which patients are assigned to treatments according to strategies derived from optimal solutions to TAB problems. We judge the ethical quality of a procedure in terms of the various criteria outlined in Section 2.3. Depending on the research setting, there may be reasons to prefer one of these criteria to another.

For the two-decision problem, the behavior of the MB is, in general, favorable. When compared with two other frequently proposed allocation rules, it exhibited the best performance on all of the ethical criteria over a majority of the parameter space. For the three-decision problem, the MB again had the best performance in general [excepting its abysmal performance with respect to $P(ND)$], but the regions over which it improves on the others is not as large - this due to the added study length enforced by the large values of the stopping parameter r.

With this proposal of the MB rule, we have worked with a sim-

ple model. While there are bona fide applications for the MB as it now exists, the areas of application would be broadened if some of the following extensions could be developed.

1. Allow a distribution on the time required to response.
2. Allow for incorporation of at least one or two prognostic factors.
3. Allow for more general distribution of the response variable.
4. Allow for different measures of discrimination between treatments.
5. Allow more than 2 treatment alternatives.

Some of these extensions (3-5) are fairly straightforward, but the others could prove to be quite difficult if they are to be made functional.

While there remain questions regarding the appropriate usage of adaptive sampling methods, it seems worthwhile to continue to study their behavior, with the goal of gradually increasing their adoption in real life. Ethical dilemmas will never disappear, but designs such as that proposed here may offer a partial easing of the conflicts.

Acknowledgment. I am grateful to Quentin Stout for algorithmic support on this project and to the referees for their helpful comments.

References

ANSCOMBE, P. (1963). Sequential medical trials. *Journal of the American Statistical Association* **58** 365-383.

BARTLETT, R., ROLOFF, D., CORNELL, R., ANDREWS, A., DILLON, P. and ZWISCHENBERGER, J. (1985). Extracorporeal circulation in neonatal respiratory failure: a prospective randomized study. *Pediatrics* **76** 479-487.

BATHER, J.A. (1985). On the allocation of treatments in sequential medical trials. *International Statistical Review* **53** 1-13.

BECHHOFER, R.E., KIEFER, J. and SOBEL, M. (1968). *Sequential Identification and Ranking Procedures,* Chicago: University of Chicago Press.

BERRY, D.A. (1978). Modified two-armed bandit strategies for certain clinical trials. *Journal of the American Statistical Association* **73** 339-345.

BERRY, D.A. and EICK, S.G. (1987). Decision analysis of randomized clinical trials: comparison with adaptive procedures (unpublished manuscript).

BERRY, D.A. and FRISTEDT, B. (1985). *Bandit Problems: Sequential Allocation of Experiments.* New York: Chapman and Hall.

BURINGER, H., MARTIN, H. and SCHRIEVER, K.H. (1980). *Nonparametric Sequential Selection Procedures.* Boston: Birkhauser.

BYAR, D.P., SIMON, R.M., FRIEDEWALD, W.T., SCHLESSELMAN, J.J., DEMETS, D.L., ELLENBERG, J.H., GAIL, M.H. and WARE, J.H. (1976). Randomized clinical trials: perspectives on some recent ideas. *New England Journal of Medicine* **293** 74-80.

CHERNOFF, H. and PETKAU, A.J. (1985). Sequential medical trials with ethical cost. *Proceedings of the Kiefer-Neyman Conference.*

FUSHIMI, M. (1973). An improved version of a Sobel- Weiss play-the-winner procedure for selecting the better of two binomial populations. *Biometrika* **60** 517-523.

GITTINS, J.C. and JONES, D.M. (1974). A dynamic allocation index for the sequential design of experiments. *Progress in Statistics* (Gani, J., ed). Amsterdam: North Holland, 241-266.

HARDWICK, J.P. (1986a). Adaptive allocation in clinical trials: The ECMO controversy, University of Michigan, Ann Arbor, MI (technical report).

HARDWICK, J.P. (1986b). The modified bandit, University of Michigan, Ann Arbor, MI (technical report).

HARDWICK, J.P. (1989). Comment: recent progress in clinical trials that adapt for ethical purposes. *Statistical Science* **4** 328-336.

HARDWICK, J. and STOUT, Q.F. (1991). Bandit strategies for ethical sequential allocation. *Proceedings of the 22nd Symposium on the Interface: Computing Science and Statistics*, 421-424.

HOEL, D.G., SOBEL, M. and WEISS, G.H. (1975). A survey of adaptive sampling for clinical trials. *Perspectives in Biometrics,* Vol. 1, New York: Academic Press, 29-61.

KIEFER, J.E. and WEISS, G.H. (1971). A truncated test for choosing the better of two binomial populations. *Journal of the American Statistical Association* **66** 867-871.

KIEFER, J.E. and WEISS, G.H. (1974). Truncated version of a play-the-winner rule for choosing the better of two binomial populations. *Journal of the American Statistical Association* **69** 807-809.

ROBBINS, H. (1952). Some aspects of the sequential design of experiments. *Bulletin of the American Math Society* **58** 527-535.

SIMON, R., WEISS, G.H. and HOEL, D.G. (1975). Sequential analysis of binomial clinical trials. *Biometrika* **62** 195-200.

THOMPSON, W.R. (1933). On the likelihood that one unknown probability exceeds another in view of the evidence of the two samples. *Biometrika* **25** 275-294.

WEI, L.J. and DURHAM, S. (1978). The randomized play the winner rule in medical trials. *Journal of the American Statistical Association* **73** 840-843.

WEINSTEIN, M.C. (1974). Allocation of subjects in medical experiments. *New England Journal of Medicine* **291** 1278-1285.

ZELEN, M. (1969). Play-the-winner rule and the controlled clinical trial. *Journal of the American Statistical Association* **64** 131-146.

DEPARTMENT OF STATISTICS
UNIVERSITY OF MICHIGAN
ANN ARBOR, MI 48109