The Use of Response-Adaptive Designs in Clinical Trials

William F. Rosenberger, PhD, and John M. Lachin, ScD

The Biostatistics Center, Department of Statistics/Computer and Information Systems, The George Washington University, Rockville, Maryland

ABSTRACT: Response-adaptive designs in clinical trials are schemes for patient assignment to treatment, the goal of which is to place more patients on the better treatment based on patient responses already accrued in the trial. While ethically attractive at first glance, these designs have had very little use in practice; yet the statistical literature is rich on this subject. We discuss procedures and properties of these designs. Particular focus is given to the randomized play-the-winner rule of Wei and Durham, which was used in the ECMO trial. We also discuss reasons for the lack of use of these models, and areas of current and future research to address the weaknesses of these methods. We conclude that these designs may be applicable in some situations and describe conditions under which such a trial may be feasible.

KEY WORDS: *ethics in clinical trials, randomized play-the-winner rule, randomization, treatment allocation*

INTRODUCTION

An adaptive design in a clinical trial incorporates emerging information from the trial into the treatment assignment probabilities in an attempt to assign more patients to the treatment performing better thus far in the trial. For example, if there are two treatments A and B, then when a patient is ready to be assigned to treatment, if treatment A appears to be more successful than treatment B thus far in the trial, that patient would have a greater than 50% chance of being assigned to treatment A. Response-adaptive designs in particular base treatment assignment probabilities on the success of the treatment with regard to the primary outcome of the trial. (Another concept, covariate-adaptive designs, will not be discussed here.) While adaptive designs have been proposed in the industrial quality control and decision theory contexts (see, for example, [1]), their application to clinical trials requires complex ethical and logistical considerations because clinical trials involve human beings—in particular, sick human beings.

Some disturbing paradoxes are encountered when considering adaptive

Address reprint requests to: William F. Rosenberger, PhD, Department of Statistics/Computer and Information Systems, George Washington University, The Biostatistics Center, 6110 Executive Boulevard, Suite 750, Rockville, Maryland 20852.

Received October 29, 1992; revised April 6, 1993.

Controlled Clinical Trials 14:471–484 (1993)

655 Avenue of the Americas, New York, New York 10010

0197-2456/93/$6.00

designs for clinical trials. Although the concept appears to be ethically attractive, particularly from a patient's point of view, the extent to which adaptive designs are truly more "ethical" than the traditional randomization designs has, to our knowledge, yet to be formally studied by ethicists. In addition, adaptive designs have generated intense criticism from logistical and statistical standpoints. Over the past 20 years, many of the leading statisticians and probabilists have done research on various aspects of adaptive designs, and yet there has been little application of the methodology. Few other scientific endeavors involving so many fine researchers have found such little practical use. In 1977, Richard Simon wrote [2]: "It is clear that the statistical literature on adaptive treatment assignment has had little impact on the conduct of clinical trials. Many of the deficiencies of current methods can be remedied with further work." Fifteen years later (a fruitful 15 years for researchers in adaptive designs) he writes [3]: "There is an immense volume of literature on [adaptive assignment] strategies yet the research has almost no applications in real clinical trials. . . . I am not optimistic about the usefulness of this area of research."

This article will address some of the practical issues involved in adaptive designs, the reasons for the sparse application of the methodologies developed, some current research areas designed to correct deficiencies, and the future of these designs. It is not intended to be an exhaustive review of the methodological and mathematical literature. Hardwick [4] gives a progress report on recent methodological developments. We begin by describing a particular adaptive design that has been used in a clinical trial; it is the randomized play-the-winner (RPW) design of Wei and Durham [5,6], a modification of Zelen's play-the-winner design [7].

THE RPW DESIGN

The RPW design can be thought of as an urn model. At the start of the trial, there are α balls of each color (say, red and black) in an urn. When a patient is available for assignment, a ball is drawn at random and *replaced*. The patient is assigned to treatment by color: if the ball is red, the patient is assigned to treatment A; if the ball is black, the patient is assigned to treatment B. When an outcome is available from the patient (it must be dichotomous: success or failure), β red balls are added to the urn if the outcome was a success on treatment A or a failure on treatment B; β black balls are added to the urn if the outcome was a success on treatment B or a failure on treatment A. In this way, the urn builds up more balls representing the more successful (or less unsuccessful) treatment, skewing the assignment probabilities accordingly. The sensitivity parameters α and β can be adjusted to suit the needs of the trial. As α is increased, the urn is diluted, making the adaptation procedure less likely to favor one treatment; as β is increased, the assignment probabilities will be skewed further toward the "better" treatment arm. This rule is denoted RPW (α, β).

If patient outcomes are not observed immediately after assignment but at some later period of observation, then balls are added to the urn when outcomes are reported. At the time of patient assignment, the urn reflects all previously reported patient outcomes. In principle, then, if a patient outcome

is not observed before the next patient is assigned to treatment, the urn model will still provide an adaptive procedure, but adaptation will take place more slowly.

THE ECMO TRIAL

The RPW design was used in a clinical trial of extracorporeal membrane oxygenation (ECMO), a surgical procedure for newborns with respiratory failure [8]. The technique had been used when infants were judged to be moribund and unresponsive to conventional treatment, which included ventilation and pharmacological therapy. Early trials on safety and efficacy indicated that the ECMO technique was safe and had an overall success rate of 56%. Historical information indicated that the conventional therapy was successful in about 20% of the cases. The RPW design was chosen because "(1) the outcome of each case [was] known soon after randomization, making it possible to use; (2) [it was] anticipated that most ECMO patients would survive and most control patients would die, so significance could be reached with a modest number of patients; [and] (3) it was a reasonable approach to the scientific/ethical dilemma" [8, p. 480]. The RPW (1, 1) rule was implemented: the first patient was assigned to ECMO and survived. The second patient was randomized to the conventional treatment and died. All remaining patients were assigned to ECMO; all survived. It was decided by ranking and selection procedures [9] that after 12 total patients had been randomized efficacy had been demonstrated; hence, the trial was stopped. Thus the RPW adaptive assignment procedure performed as intended: most patients were assigned to the treatment that appeared to be superior.

Serious questions and some complex issues arose as a result of this trial. The foremost question is how can two treatments be adequately compared when only one patient was assigned to one of the treatments? In retrospect, the investigators note that, had they used $\alpha > 1$, more babies would have been randomized to the conventional treatment and the trial would have been more conclusive [8,10] (but more babies may have died). Also, the results of any trial with a total sample size as small as 12 would be suspect, even if equal allocation were used. In addition, physicians and nonstatisticians in the medical community are accustomed to using a P value to determine efficacy in a trial. They are unfamiliar with the ranking and selection procedures used in the ECMO trial.

Wei addresses this issue by developing an exact permutation test to draw inferences from an RPW trial [11]. Using a networking algorithm, he finds the exact distribution of the number of successes on treatment A under a permutation model (which assumes that patient response is deterministic). However, Begg believes that Wei's approach is inappropriate because the sample space defined by Wei's analysis includes the case where all patients were randomized to ECMO, and hence there would be no information about the treatment comparison [12]. He presents other ways of analyzing the data, conditioning on different margins, and arrives at P values ranging from .038 to .62. Clearly, analysis issues are as much an enigma as the design issues. Pocock [13] believes the issue to be moot. He states that

> my use of this ECMO example in teaching this year has on both occasions resulted in widespread amusement by the course participants. Such ridicule of the 0/11 versus 1/1 response comparison by both statisticians and post-graduate medical students is to my mind a healthier reaction to weird and inadequate data than formalized debate over what is the correct P value.

Some have questioned the ethics of performing a clinical trial when historical information so strongly favored ECMO over the conventional treatment [14]. However, the primary investigators of the trial point out that ECMO is an invasive, complicated, and costly procedure, and ECMO was not universally available at the time of the trial, even in some well-known clinics [15].

One other ethical issue has been raised: Zelen's postrandomization consent rule [16] was employed, in which only patients assigned to the experimental treatment are asked for consent, so that only parents of babies randomized to ECMO were notified about the trial and the two available therapies. This meant that the parents of the one baby randomized to the conventional treatment were not informed about the ECMO procedure; that baby died.

HISTORY OF ADAPTIVE DESIGNS

As early as 1969, Cornfield, Halperin, and Greenhouse discussed the ethical issues regarding adaptive designs [17]:

> Application of these results might ease the ethical problem involved in trials on human subjects. The usual ethical justification for not administering an agent of possible efficacy to all patients is the absence of definite information about its effectiveness. However satisfactory this justification may be before the trial starts it rapidly loses cogency as evidence for or against the agent accumulates during the course of the trial. But any solution . . . which permits adaptive behavior . . . at least reduces this ethical problem.

Later, Weinstein's 1974 special article in the *New England Journal of Medicine* [18] strongly advocated adaptive allocation as an alternative to traditional assignment rubrics, taking the view that

> any decision rule for allocating patients to clinical procedures in any way other than according to the best interest of the subject at hand does entail a sacrifice on the part of the subject

and concluding with the opinion that

> adaptive methods should be used as a matter of course. It never pays to commit oneself to a protocol under which information available before the study or obtained during its course is ignored in the treatment of a patient.

This led to a flurry of commentary on the subject (see [19]).

Zelen's play-the-winner and Wei and Durham's RPW designs evolved from the two-armed bandit problem in gambling, which was first studied in the experimental design context by Robbins [20]. Other adaptive designs evolved along the lines of two-stage sequential trials [21,22], where 50–50 allocation was employed until there was clear evidence that one treatment was more

efficacious, and then *all* remaining patients were assigned to the better treatment. Ware used this design for another trial of ECMO [23,24] following the unconvincing RPW trial of ECMO. Two-stage designs were later followed by multistage designs [17], where, at some given points in the randomization process, a new stage is started with all existing information used to update a prior distribution on the treatment difference. Such a procedure would require large samples and could be relevant in the context of today's complex multicenter trials.

Robbins and Siegmund [25] developed a sequential probability ratio test that minimizes the expected number of patients assigned to the inferior treatment. Other authors [26–28] give cogent summaries of sequential inference procedures, in conjunction with adaptive assignment schemes, with no fixed patient horizons. We do not consider those designs involving open-ended sequential inference in this article.

STATISTICAL ISSUES IN ADAPTIVE DESIGNS

We now summarize some important statistical issues in adaptive designs, many of which remain unresolved or unmentioned in the literature. We review recent research addressing these issues and concerns.

Inference

Adaptive designs are only useful if there are adequate inferential techniques to accurately determine the degree of treatment effect by some criterion at the conclusion of the trial. Simon [2] states that "any design that does not provide for a convincing test of the null hypothesis has little chance of being adopted for general use." The controversy over the appropriate analysis of the ECMO data is evidence that the appropriate inferential procedures in small scale RPW trials has not yet been resolved. Wei et al. [29] developed small-sample confidence limits and a large-sample test statistic under a simple population model where the probability of successes on treatments A and B are assumed constants. That is, they assume that the sequence of patients who arrive for entry into the trial represents samples drawn at random from two homogeneous populations, with no drift in the probabilities of success. They show that the usual test of the difference of two proportions is asymptotically normal when taking the RPW allocation procedure into consideration. Rosenberger [30] found an upper bound on the rate of convergence to normality of the test statistic, which is slower than in the usual coin-tossing randomization case. Because of the complex dependent structure of the sequence of treatment assignments, a larger sample size must be employed for the normal approximation to hold than would be necessary under simple randomization.

Rosenberger [31] also developed a large sample test statistic that can be thought of as an asymptotic version of Wei's permutation test statistic [11], which is asymptotically normal under certain conditions on the sequence of responses. Simulation shows that the test is conservative for $\alpha = 1$ and a good approximation to normality is obtained for $\alpha = 5$.

In this age of computing, finding asymptotic distributions of test stat-

istics for adaptive designs may not be necessary, as one could simulate the distribution of any test statistic under a specified population model. In principle, one can use simulation to evaluate whether or not the test statistic has the nominal size. Another option would be to approximate the sampling distribution of the test statistic using a bootstrap technique. Therefore, inference should not be an obstacle to the use of adaptive designs. However, either approach would entail extensive programming by the clinical trial statisticians, who may already be excessively busy in study design, management, and analysis.

For those inferential procedures developed which assume that the patients who arrive for entry into the trial represent a random sequence of patients from a homogeneous population, a criticism has been made that the characteristics of patients entering later in the trial may differ substantially from those patients entering earlier [19]. In many trials (an example is cited in [19]), the survival experience of patients recruited early in the trial differs substantially from that of patients recruited later. Simon et al. examine the effects of heterogeneous patient characteristics via simulation and show that when such experimental bias exists the α level of the test is greatly inflated [32]. One of the principal issues to be addressed, then, is how best to plan for the conduct of a trial such that there is no "drift" in the patient characteristics or risk factors.

Polychotomous or Continuous Outcomes

The RPW rule is applicable only for dichotomous outcomes. Many trials have as their outcome measure some continuous or polychotomous measurement (e.g., a trial on hypertension with blood pressure as the outcome). An assignment procedure has been proposed for an adaptive assignment scheme for polychotomous or continuous outcomes [31]. The idea involves calculating a standardized nonparametric rank test at each stage in the trial to establish the assignment probabilities. While the mathematical considerations for the analysis of this design are messy, the randomization procedure could be implemented by computer. Simulations show that the test statistic given is close to normal for moderate to large sample sizes (>50). This procedure is not computationally feasible using the techniques of Wei [11] in the small-sample setting.

Secondary Outcomes and Covariates

Clinical trials are often not only concerned with single outcomes but also with secondary outcomes. For example, a factor, such as toxicity, may be an important issue in addition to the effect of the treatment on efficacy. It is not clear how to take secondary outcomes into consideration in the adaptation process. The analysis of a secondary outcome when adaptation was performed with respect to only the primary outcome is not clear either. This is a possible area for future research.

Another important consideration in the analysis of data from clinical trials is the effect of baseline covariates, such as patient characteristics, on the outcome of interest. Some work has been done under a Bayesian framework

by Woodroofe [33,34] and Sarkar [35]; however, it is not clear how to proceed with baseline covariate adjustment and modeling under an RPW-type assignment scheme.

Procedures that adjust the assessment of treatment effects for other covariates could be employed when it is found that the sequence of patients entered into the trial shows a time trend with respect to such covariates. This could in part solve the problem of "drift" in patient characteristics or risk factors outlined in the section "Inference."

Immediate Outcomes

For many trials (such as cancer trials) few outcomes are observed until all or most patients have been treated. Almost all trials have some delay in outcomes, not only in observing but also in reporting. While the RPW design does not require instantaneous outcomes, in many trials the outcomes are not immediate enough to obtain any benefit from adaptation. Many outcomes concern time to some event, and often patient entry is staggered. The procedures discussed so far would not be relevant in this situation. Louis [36] and Flehinger and Louis [37] developed sequential designs based on exponential survival models. A biased coin approach might also be possible, using as the assignment probability some function of the current Kaplan–Meier estimates of the two treatments or the current values of a linear rank test at the time of the assignment. However, the Kaplan–Meier estimate is known to be sensitive in the tails, so a risk ratio estimate may be a more appropriate measure of treatment efficacy.

Several methods could be employed to incorporate delayed outcomes into an urn design such as the RPW design. One solution might be to model the patient arrivals for assignment and the patient outcomes as a queuing process with some distributional assumption on the arrival and outcome times. One could also employ a group sequential-type plan, where one adapts only after a group of patients have outcomes available, rather than adapting after each individual outcome. Another alternative is to adapt on the basis of some surrogate measure. For example, in a trial on AIDS, the outcome may be mortality, but one could adapt on the basis of CD4 counts, which could be ascertained more quickly.

Comparison of More than Two Treatments

Some clinical trials are designed to compare more than one treatment or several dose levels of a single treatment. Generalizations to existing methods would have to be developed to account for these types of trials. In terms of the comparison of several treatments with Bernoulli responses, Bather and Coad [38] developed a sequential procedure to eliminate the inferior treatments early for ethical purposes. Whether this design could be generalized to the response-adaptive setting is a topic for further research. In terms of dose-response designs, Durham and Flournoy [39] considered random walk rules which sequentially assign treatment levels to subjects, adapting to avoid assigning toxic doses.

Interim Monitoring of Data

Another important need besides that of a test statistic is an appropriate procedure for interim monitoring during the course of the clinical trial. Most clinical trials with a fixed patient horizon require some form of interim monitoring, and a group sequential design is essential. It is not clear how to account for group sequential monitoring of data in an adaptive design framework.

Sample Size and Power

Finally, in most (nonsequential) clinical trials, it is important to perform a preliminary calculation of the adequacy of the proposed sample size in terms of the power of the principal test of the primary outcome or in terms of the precision of the estimate of the critical difference between groups for which the outcome is to be assessed. Both involve the concept of some underlying population parameters that apply to the representative patient population. Much of the past work on statistical inference for these designs has focused on the conduct of statistical tests under the null hypothesis. The distribution of these tests under alternative hypotheses has yet to be fully explored. Thus, at the present time, it would not be possible to perform a sample size evaluation for a clinical trial that would be based on the distribution of the appropriate test statistic under an alternative of interest.

Although power is often considered to be a secondary consideration to the ethical advantages gained by using an adaptive design with an associated statistical test, the need to accurately report nil results (i.e., no treatment difference) makes power considerations important in any clinical trial. The imbalance in treatment assignments from an adaptive design may reduce power. However, in many cases (Lachin shows this for a normal deviate test [40]) allocation ratios must exceed 7:3 in order to severely impact power.

This discussion is moot if one is considering ranking and selection procedures or a sequential adaptive design with no fixed patient horizon. However, in the context of sequential designs, one must be concerned about balancing the probabilities of type I and type II errors along with the ethical concerns of minimizing the total sample size and the expected number of patients on the inferior treatment (see, for example, a sequential clinical trial design by Samuel-Cahn and Wax [41], where the probability of a type I error had to be set at 0.30 to take into account ethical considerations in the trial).

ISSUES CONCERNING IMPLEMENTATION OF ADAPTIVE DESIGNS

Many clinical trials today are designed as multicenter studies to ease problems associated with recruiting adequate numbers of patients. Because adaptive designs have had limited use, none to our knowledge in multicenter settings, many issues with respect to implementation are unresolved.

Complexity and Logistical Concerns

Clinical trials, to all involved, can be logistical nightmares. There must be close cooperation between the physicians, the statisticians, the clinicians, the data personnel, and the patients for a trial to run smoothly. Any complexity

in the design of a trial can lead to an increase in mistakes in assigning treatments to patients and in executing the trial. Pocock states [42] that "any assignment procedure must be simple, rapid, objective and foolproof and these conditions are hard to fulfil with an adaptive scheme." Clearly, these adaptive schemes are best implemented through an automated procedure. This would especially be required in a multicenter trial. Some consideration should also be given to lags in reporting of patient responses in this case.

Stratification

In many clinical trials, some form of stratification is employed. If the reason for stratification is to allow for heterogeneity of treatment effect among strata, then some form of an adaptive procedure specific to the patient responses within each stratum might be appropriate. Considerations of the appropriate method of analysis following a stratified adaptive allocation procedure have yet to be addressed.

The Ethics of Conducting a Trial When Information Favors One Treatment

Some would argue that if there is enough evidence to justify assigning one treatment with greater probability than another, then one must ethically terminate the trial and assign all patients to the seemingly better treatment. For instance, Royall writes [14] that "the ethical problems are clear: after finding enough evidence favoring A to require reducing the probability of B, the physician . . . must see that the next patient gets A, not just with high probability, but with certainty." It should be noted (see [43]), however, that interim results in a clinical trial do not necessarily provide evidence of treatment efficacy, even if one of the treatment arms is doing better. Simon states [3] that he "does not find it attractive to approach a patient saying that I do not know which treatment is better, but treatment A is doing better therefore I will give you a greater than 50 percent chance of getting it." However, we assume that, from a patient's perspective, having more of a chance of being randomized to the treatment doing better in the trial so far, regardless of any conclusive efficacy, should be preferable to a 50–50 randomization, provided that the trial will yield valid and convincing information. Perhaps this presumption should be further studied in its own right by population ethicists and behaviorists.

Masking of Assignments

One of the issues raised in the context of adaptive designs is whether or not the patient and physician should be made aware of the current trend at the time of assignment and whether the participating physicians should be aware of the treatment assignments as they are provided. If the trial is double-masked, then the treatment assignment procedure will be free of a variety of treatment selection biases that might arise in the selection of order of entry of patients into the trial, and would be free of various observer biases in the

evaluation of the outcomes of treatment [40]. However, subtle biases may arise from the patient's or the physician's perspective if the treatment assignment probabilities or if the actual treatment assignments are unmasked. The susceptibility of these designs to these types of selection and observer biases requires further consideration.

Another class of bias concerns an experimental bias related to differences in the distribution of treatment groups with respect to important covariates. The susceptibility of these designs to such biases also requires further consideration.

Consent

An important legal and ethical aspect of clinical trials concerns what the patient is told during the process of consent. (The problems raised in the ECMO trial with respect to consent are clear from an earlier discussion.) If the treatment assignment is to be masked, then the patient need only be told that one of two (or more) treatments will be administered and that the probability of treatment assignment will depend on the relative frequency of favorable and unfavorable responses to treatments among those patients previously entered into the trial. We do not feel it is appropriate or necessary that the patient be told the actual prior trend of patient responses or to be told the probabilities of assignment to the various treatments. Again, disclosure of the probability of treatment assignment is tantamount to disclosure of the current trend of the data.

CONDITIONS CONDUCIVE TO ADAPTIVE DESIGNS TODAY

The question arises, given all these statistical and logistical concerns, are there conditions under which an adaptive design could be implemented successfully given the existing methodology? We now outline some general conditions under which we believe that an adaptive design may be feasible:

1. There is a single outcome or hypothesis of interest.
2. Outcomes are ascertainable in a short period of time.
3. The study has important public health consequences, but the diseases studied are not life threatening.
4. The study has an adequate sample size and the composition of the sample is not likely to change over time.
5. The participants in the study have the resources to logistically implement the design.

Under such conditions, an urn design such as the RPW procedure for dichotomous outcomes, or Rosenberger's modified version for polychotomous or continuous outcomes could be used. Regardless of the design employed, a provision should be made so that a specified minimal number of patients is assigned to either treatment arm.

Such conditions are met in some areas of medical research. One example might be an evaluation of medical therapy to control hypotension or hypertension following medical procedures known to affect blood pressure (such as hypotension during regional anesthesia).

DISCUSSION

If it were possible to apply an adaptive design that yields substantial gains ethically, without any loss of scientific rigor or credibility compared to the traditional randomization design, then surely such designs would become common. The successful adaptive trial must have a sample size sufficiently large and enough patients must be assigned to both treatments so that a convincing comparison of the two treatments can be made. D. R. Cox says that "there is an imperative to establish minimum sample sizes" for RPW designs [44]. A convincing test of the null hypothesis must be available. Results by Wei et al. for dichotomous outcomes [29] and Rosenberger for dichotomous or polychotomous outcomes [31] seem promising for large-sample trials, depending on whether a permutation or population model is chosen. However, methods must be developed that allow for appropriate analysis of secondary outcomes and covariate adjustment. It is yet unclear as to how to adapt with survival data.

It might be argued that circumspect statisticians should try these designs in trials with outcomes that are not grave. This may be wise to "work out the bugs," but it should be remembered that the full ethical benefit of an adaptive design would not be realized unless the therapies studied are of potentially great consequence to public health.

One might argue in retrospect that any trial involving the mortality of babies as an outcome opens the door to intense ethical concerns and emotions that would not be encountered in a study of consenting adults suffering with a more benign condition. There are indeed problems with adaptive designs, but should they be abandoned on the basis of a highly atypical and unsuccessful trial like ECMO?

Are there trials where an adaptive design is ethically and logistically feasible? Would the statistical community be bold enough to support such a design? We believe that any ethical gain must be weighed against the practicality of executing the trial. If an adaptive trial could be conducted with no obvious losses in validity or amount of information gained, then ethical considerations might demand their use. However, it is clear that many problems are left to be solved before these conditions will be satisfied.

Most clinical trials are, in a broad sense, adaptive. Patients are allocated to two or more different treatments until one is found to be more effective; the future patients will be assigned to the treatment declared superior. But when should adapting begin? Should we wait until a conclusive hypothesis is rejected, or can we begin to adapt as data emerge; i.e., can we "play-the-leader" instead of "playing-the winner?"

It is the statistician's job to explain, justify, and defend complex designs and analysis of the results to clinicians and others involved in the trial. Commenting on this, Bailar writes [45]: "The clinicians and statisticians who really control decisions on study designs have not been convinced that these elegant approaches offer any real advantage over plain old-fashioned randomization which even physicians have come to accept, trust and sometimes demand."

Like others, we conclude that the future usefulness of adaptive designs is uncertain. Armitage's comment [28] that "none of [the negative] points taken singly seems to me to argue strongly against [adaptive] allocation, although

taken together with the . . . ethical problems, . . . they present a rather discouraging picture" is particularly relevant. Perhaps one explanation for this skepticism regarding adaptive designs is that some of the research has been carried out by theorists who have little experience in the "real world" clinical trials setting. Also, the failure of ECMO has perhaps discouraged clinical trial statisticians from taking the risk to propose an adaptive design, even in settings that might be appropriate. Much of the criticism of adaptive designs derives from their lack of use in actual clinical trials. It is precisely this catch-22 situation that holds the future of these designs in limbo.

It is perhaps instructive to note that, as demonstrated at a recent conference on adaptive designs sponsored by the American Mathematical Society and the Institute of Mathematical Statistics, adaptive designs for clinical trials is a dynamic area for research; many of the generalizations in the methodology needed and problems to be overcome with adaptive designs discussed in this article present challenging mathematical problems which could permeate the literature and doctoral theses for years to come. It is perhaps even more instructive to note, however, that very few practical clinical trialists attended the conference.

Given time, the solutions to many of the questions addressed in this article may become apparent. It would be nice to think that if in the future the perfect trial for an adaptive design comes along, we will have the methodology available to successfully carry it out and analyze the results.

The authors thank Prof. R. T. Smythe for helpful comments on an earlier version of this manuscript and for sharing his insights into these interesting problems with the first author over the past 4 years. Some of the ideas presented here are products of discussions with Prof. Stephen D. Durham and Prof. Nancy Flournoy; others were suggested by the anonymous referees of this paper. Parts of this research were presented by the first author at the 1992 AMS Summer Conference on Adaptive Designs with funding from the National Science Foundation, and in other colloquia. Some of the thoughts in this paper arose from discussions following these talks.

REFERENCES

1. Schneider H, Tang K: Adaptive procedures for the two-stage group-testing problem based on prior distributions and costs. Technometrics 32:397–405, 1990
2. Simon R: Adaptive treatment assignment methods and clinical trials. Biometrics 33:743–749, 1977
3. Simon R: A decade of progress in statistical methodology for clinical trials. Stat Med 10:1789–1817, 1991
4. Hardwick J: Comment: recent progress in clinical trial designs that adapt for ethical purposes. Stat Sci 4:327–336, 1989
5. Wei LJ, Durham S: The randomized play-the-winner rule in medical trials. J Am Stat Assoc 73:840–843, 1978
6. Wei LJ: The generalized Pólya urn design for sequential medical trials. Ann Stat 7:291–296, 1979
7. Zelen M: Play the winner rule and the controlled clinical trial. J Am Stat Assoc 64:131–146, 1969
8. Bartlett RH, Roloff DW, Cornell RG, Andrews AF, Dillon PW, Zwischenberger JB: Extracorporeal circulation in neonatal respiratory failure: a prospective randomized study. Pediatrics 76:479–487, 1985

9. Cornell RG, Landenberger BD, Bartlett RH: Randomized play the winner clinical trials. Commun Stat—Theor Meth 15:159–178, 1986
10. Wei LJ: Discussion of paper by C. B. Begg. Biometrika 77:476–477, 1990
11. Wei LJ: Exact two-sample permutation tests based on the randomized play-the-winner rule. Biometrika 75:603–606, 1988
12. Begg CB: On inferences from Wei's biased coin design for clinical trials. Biometrika 77:467–484, 1990
13. Pocock SJ: Discussion of paper by C. B. Begg. Biometrika 77:480–481, 1990
14. Royall RM: Ethics and statistics in randomized clinical trials. Stat Sci 6:52–62, 1991
15. Bartlett RH, Cornell RG: Comment. Stat Sci 5:63–65, 1991
16. Zelen M: A new design for randomized clinical trials. N Engl J Med 300:1242, 1979
17. Cornfield J, Halperin M, Greenhouse SW: An adaptive procedure for sequential clinical trials. J Am Stat Assoc 64:759–770, 1969
18. Weinstein MC: Allocation of subjects in medical experiments. N Engl J Med 291:1278–1285, 1974
19. Byar DP, Simon RM, Friedewald WT, Schlesselman JJ, DeMets DL, Ellenberg JH, Gail MH, Ware JH: Randomized clinical trials: perspectives on some recent ideas. N Engl J Med 295:74–80, 1976
20. Robbins H: Some aspects of the sequential design of experiments. Bull Am Math Soc 58:527–535, 1952
21. Anscombe FJ: Sequential medical trials. J Am Stat Assoc 58:365–384, 1963
22. Colton T: A model for selecting one of two medical treatments. J Am Stat Assoc 58:388–401, 1963
23. O'Rourke PP, Crone RK, Vacanti JP, Ware JH, Lillehei CW, Parad RB, Epstein MF: Extracorporeal membrane oxygenation and conventional medical therapy in neonates with persistent pulmonary hypertension of the newborn: a prospective randomized study. Pediatrics 84:957–963, 1989
24. Ware JH: Investigating therapies of potentially great benefit: ECMO. Stat Sci 4:298–340, 1989
25. Robbins H, Siegmund DO: Sequential tests involving two populations. J Am Stat Assoc 69:132–139, 1974
26. Hoel DG, Sobel M, Weiss GH: A survey of adaptive sampling for clinical trials. In Perspectives in Biometrics. Elashoff RM (Ed). New York, Academic Press, 1975, pp. 29–61
27. Bather JA: On the allocation of treatments in sequential medical trials. Int Stat Rev 53:1–13, 1985
28. Armitage P: The search for optimality in clinical trials. Int Stat Rev 53:15–24, 1985
29. Wei LJ, Smythe RT, Lin DY, Park TS: Statistical inference with data-dependent treatment allocation rules. J Am Stat Assoc 85:156–162, 1990
30. Rosenberger WF: Asymptotic Inference Problems Arising from Clinical Trials Using Response-Adaptive Treatment Allocation. Graduate School of Arts and Sciences, George Washington University, 1992, doctoral dissertation
31. Rosenberger WF: Asymptotic inference with response-adaptive treatment allocation designs, Ann Stat, 21, 1993
32. Simon R, Hoel DG, Weiss GH: The use of covariate information in the sequential analysis of dichotomous response experiments. Commun Stat—Theor Meth 6:777–788, 1977
33. Woodroofe MB: A one-armed bandit problem with a concomitant variable. J Am Stat Assoc 74:799–806, 1979
34. Woodroofe MB: Sequential allocation with covariates. Sankhya A 44:403–414, 1982
35. Sarkar J: One-armed bandit problems with covariates. Ann Stat 19:1978–2002, 1991

36. Louis TA: Sequential allocation in clinical trials comparing two exponential survival curves. Biometrics 33:627–634, 1977
37. Flehinger BJ, Louis TA: Sequential treatment allocation in clinical trials. Biometrika 58:419–426, 1971
38. Bather JA, Coad DS: Sequential procedures for comparing several medical treatments. Commun Statist—Seq Anal 11:339–376, 1992
39. Durham SD, Flournoy N: Random walks for quantile estimation. To appear in Berger J, Gupta S, eds., Statistical Decision Theory and Related Topics V. New York, Springer-Verlag, 1994
40. Lachin JM: Statistical properties of randomization in clinical trials. Controlled Clin Trials 9:289–311, 1988
41. Samuel-Cahn E, Wax Y: A sequential test for comparing two infection rates in a randomized clinical trial, and incorporation of data accumulated after stopping. Biometrics 42:99–108, 1986
42. Pocock SJ: Allocation of patients to treatments in clinical trials. Biometrics 35:183–197, 1979
43. Byar DP: Comment. Stat Sci 6:65–68, 1991
44. Cox DR: Discussion of paper by C. B. Begg. Biometrika 77:483–484, 1990
45. Bailar JC: Patient assignment algorithms—an overview. In Proceedings of the 9th International Biometric Conference, Vol. I. Raleigh, Biometric Society, 1976, pp. 189–206