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

ADAPTIVE DESIGNS IN CLINICAL TRIALS: SOME ISSUES WITH EMPHASIS ON ASYMPTOTIC INFERENCE

BY WILLIAM F. ROSENBERGER

The George Washington University

Abstract

Response-adaptive designs in clinical trials involve incorporating accruing information from patient responses to treatment into the randomization scheme in order to assign more patients to the treatment performing better thus far in the trial. While ethically attractive at first glance, these designs are rarely used in practice. A brief overview of the ethical and logistical concerns for this apparent paradox are given, focusing in particular on the randomized play-the-winner design of Wei and Durham. Some asymptotic results are then presented for adaptive strategies. A large-sample permutation test statistic is derived for the randomized play-the-winner design. The martingale central limit theorem is employed to show asymptotic normality under certain conditions on the sequence of responses. A new adaptive design is then proposed to handle cases where the responses are polychotomous or continuous. A large-sample test statistic is given. Although a rigorous proof of asymptotic normality has eluded us, simulation evidence is presented which strongly indicates asymptotic normality.

Received September 1992; revised April 1993.
AMS 1991 subject classifications. Primary 62E20; secondary 62G10.
Key words and phrases. Martingale central limit theorem, permutation test, randomized play-the-winner rule.

1. Issues in response-adaptive designs for clinical trials. Response-adaptive designs in clinical trials incorporate accruing information from patient responses to treatment into the treatment assignment probabilities in an attempt to randomize more patients to the treatment performing better thus far in the trial. While adaptive designs have been proposed in other contexts, 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 designs for clinical trials. First, although the concept appears to be ethically attractive, particularly from a patient's point of view, adaptive designs have generated intense criticism from both an ethical and logistical standpoint. Second, over the past 20 years, many of the leading scientists in the field have done research involving 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 "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" [Simon (1977)]. Fourteen years later (a fruitful 14 years for researchers in adaptive designs), he writes "There is an immense 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" [Simon (1991)].

One example of a response-adaptive design which has been used in an actual clinical trial is the randomized play-the-winner (RPW) design of Wei and Durham (1978). 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 to one of two treatments (say A and B), a ball is drawn at random and replaced. If the ball drawn was red, the patient is assigned to treatment A; if the ball was black, the patient is assigned to treatment B. When a response is available from the patient (it must be dichotomous: success or failure), β red balls are added to the urn if the response was a succession treatment A or a failure on treatment B; β black balls are added to the urn if the response was a success on treatment B or a failure on treatment A. It is known that this design tends to assign more patients to the treatment performing "better" [i.e., more successes or less failures; see Wei (1979)]. The parameters α

and β can be adjusted to skew the assignment probabilities to suit the needs of the trial. This rule is denoted RPW(α, β).

An RPW design was used in a clinical trial of extracorporeal circulation oxygenation (ECMO) versus a conventional treatment on newborns with respiratory failure. For details of this trial, see Bartlett, Roloff, Cornell, Andrews, Dillon and Zwischenberger (1985). The RPW(1,1) rule was implemented: the first patient was assigned to ECMO and survived. The second patient was assigned to the conventional treatment and died. All remaining patients were assigned to ECMO; all survived. It was decided by ranking and selection procedures [Cornell, Landenberger, and Bartlett (1986)] that, after 12 patients had been randomized, efficacy had been demonstrated, and the trial was terminated. Serious questions arose as a result of this trial. Royall (1991) gives a synopsis of the ensuing controversy. The foremost question is how can two treatments be adequately compared when only one patient was assigned to one of the treatments? Also, the results of any trial with a total sample size of 12 would be suspect. In addition, the medical community is 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. To address this last issue, Wei (1988) developed an exact permutation test to draw inferences from an RPW trial. It is apparent from the lively debate sparked by this article [see Begg (1990) with commentary] that there is no consensus on how to analyze these types of trials. Wei, Smythe, Lin and Park (1990) also present an exact and large sample inferential procedure under a simple population model.

It is doubtful at this time that adaptive randomization schemes will be used in future clinical trials unless a number of criticisms are addressed by the statistical community. First, "any assignment procedure must be simple, rapid, objective, and foolproof" [Pocock (1979)]. Second, while a design might be mathematically beautiful, clinical trials are dynamic studies involving many uncertainties. An adaptive design must be itself "adaptable," in the sense that it must be amenable to changes precipitated by ethical, logistical or analytical concerns. Rigid decision rules might have to be overruled by common sense. Third, there must be a sufficient sample size and sufficient patient assignment to both treatments to establish convincing results. Fourth, there must be appropriate inferential procedures to apply at the conclusion of a clinical trial to provide for a convincing test of the null hypothesis. Rosenberger and Lachin (1993) give a thorough summary of some of

the other issues in the practical application of these designs, and give some criteria under which adaptive designs could be employed.

With regard to inference, Rosenberger (1993) has developed large-sample permutation tests that could be used to test the hypothesis of no treatment effect under some adaptive strategies. We now summarize these results. Section 2 presents a large-sample test statistics that can be thought of as a large-sample analog of Wei, Smythe, Lin and Park's (1990) procedure. Section 3 presents a new adaptive randomization scheme for a generalization to polychotomous or continuous outcomes, along with a similar test statistic.

2. A permutation statistic for dichotomous outcomes. Let $Y_1, \ldots, Y_n$ be a sequence of dichotomous treatment assignments, where $Y_j = 1$ or 0 according to whether patient j is assigned to treatment A or B, respectively. Let $x_1, \ldots, x_n$ be a sequence of dichotomous outcomes, treated here as deterministic (since, under the null hypothesis, a patient's outcome does not depend on the treatment assigned) that take the value 1 if the treatment is successful and 0 if not. It will be assumed here, for simplicity, that responses are instantaneous. Define $z_j = 2x_j - 1$, $j = 1, \ldots, n$ (i.e., z_j takes the value 1 for a success and -1 for a failure). Let $\Im_j \equiv \sigma(Y_1, \ldots, Y_j)$, $j = 1, \ldots, n$, be the sigma algebra generated by the first j treatment assignments, and let $\Im_o$ be the trivial sigma algebra. For the rule RPW$(\alpha, 1)$, it is easily seen that the conditional probability $p_i \equiv P(Y_i = 1 \mid \Im_{i-1})$ is given by

$$\begin{aligned} p_1 &= 1/2, \\ p_i &= \frac{\alpha + S_{i-1}}{2\alpha + i - 1}, \ i > 1, \end{aligned}$$

where

$$S_i = \sum_{j=1}^{i} \left\{ z_j \left(Y_j - \frac{1}{2} \right) + \frac{1}{2} \right\}, \ i = 1, \ \ldots, \ n.$$

The test statistic of interest has numerator

$$\sum_{j=1}^{i} \left\{ z_j \left(Y_j - \frac{1}{2} \right) \right\},$$

significantly large absolute values of which will lead to rejection of the hypothesis of equal treatment effects. The key step in a proof of asymptotic normality is to equate this statistic, suitably normalized, to a certain martingale. To this end, let $\{b_{jn}\}$, $j = 1, ..., n$, be a deterministic triangular array, chosen to make

$$\sum_{j=1}^{n} z_j \left(Y_j - \frac{1}{2}\right) = \sum_{j=1}^{n} b_{jn} z_j \left(Y_j - p_j\right)$$

for each n. This choice of b_{jn} gives the equivalence of

$$T_n = 2 \frac{\sum_{j=1}^{n} z_j \left(Y_j - \frac{1}{2}\right)}{\left(\sum_{j=1}^{n} b_{jn}^2\right)^{\frac{1}{2}}}$$

and

$$W_{nn} = 2 \frac{\sum_{j=1}^{n} b_{jn} z_j \left(Y_j - p_j\right)}{\left(\sum_{j=1}^{n} b_{jn}^2\right)^{\frac{1}{2}}},$$

which is the nth term of a martingale difference array. It is not difficult to verify that the desired sequence $\{b_{jn}\}$ is given by

(2.1)

$$\begin{aligned} b_{nn} &= 1, \\ b_{jn} &= \prod_{k=j+1}^{n} \left(1 + \frac{z_k}{2\alpha + k - 1}\right), j = 1, \ldots, n-1. \end{aligned}$$

By Corollary 3.1 of Hall and Heyde (1980, pp. 58-59), T_n converges in distribution to a standard normal variate under the following two conditions:

(2.2)
$$\max_{1 \le j \le n} \frac{b_{jn}^2}{\sum_{k=1}^{n} b_{kn}^2} \to 0 \text{ as } n \to \infty.$$

and

(2.3)
$$S_n/n \to 1/2 \text{ in probability as } n \to \infty.$$

Because we are treating the responses $\{x_i\}$ as a deterministic sequence, one could never discern from a finite sample whether (2.2) and (2.3) will hold. Indeed, at one extreme when all responses are successes (all $x_i = 1$), we have a Pólya urn model, and neither condition holds; in this case, S_n/n has a beta limit [see Athreya and Ney (1972) p. 220]. At the other extreme when all are failures (all $x_i = 0$), we have Bernard Friedman's urn [see Friedman (1949)], and asymptotic normality follows from Freedman (1965).

To determine how a "typical" response sequence might behave, consider the $\{x_i\}$ to be realizations of a sample $X_1, \ldots, X_n$ from a sequence of Bernoulli trials with parameter p. Let $Z_i \equiv 2X_i - 1$, $i = 1, \ldots, n$, and let the random variables $\{B_{jn}\}$ be defined via (2.1) with $\{z_j\}$ replaced by $\{Z_j\}$. We now state two theorems, proofs of which can be found in Rosenberger (1993).

THEOREM 2.1. *When the $\{x_i\}$ are realizations of independent Bernoulli trials with success probability p, then*

$$var\ (S_n/n) \to 0 \text{ as } n \to \infty \text{ for any } p < 1,$$

where the expectation is taken over the $\{X_i\}$ and the $\{Y_i\}$.

THEOREM 2.2. *Under the assumptions of Theorem 2.1, if $p <$* 0.75,

$$\max_{l \le j \le n} \frac{B_{jn}^2}{\sum_{k=1}^{n} B_{kn}^2} \to 0 \textit{in probability as } n \to \infty.$$

REMARK. It is easy to show that Theorems 2.1 and 2.2 remain valid for somewhat more general sequences $\{X_i\}$. If the $\{X_i\}$ are an independent Bernoulli sequence with $P\{X_i = 1\} = p_i$, Theorem 2.1 holds, provided that, for some $p < 1$, $p_i \to p$ as $i \to \infty$; Theorem 2.2 holds, if $p_i \le p \le 0.75$ for all i.

Thus for $p < 0.75$ and random sequences of $\{X_i\}$ generated in the fashion described above, the two conditions of the martingale central limit theorem are satisfied. Ideally, we would like to prove that, for almost all sequences generated in this fashion, T_n has a central limit. However, based on the results of these theorems, the most that can be stated is given in the following corollary.

COROLLARY 2.1. *For $p < 0.75$ and almost all realizations of random sequences $\{X_i\}$ generated as Bernoulli random variables with parameter p, there exists a subsequence n_k such that $T_{n_k} \to N(0,1)$ in law.*

Extensive simulations were conducted to examine the behavior of the test statistic under the Bernoulli assumption. Details are given in Rosenberger (1992). The speed convergence in conditions (2.3) and (2.2), not surprisingly, depends heavily on the value of p, and slows considerably as p exceeds 0.75. Coverage probabilities were calculated and compared to the tails of the normal distribution for $\alpha = 1$ and $\alpha = 5$ and $n = 30, 50$, and 100. Responses were generated under the Bernoulli assumption for various values of p. For $\alpha = 1$, the test statistic is conservative, very much so in the extreme tails. For $\alpha = 5$ the results are somewhat conservative in the extreme tails, but even for $n = 30$ and $n = 50$, coverage is close to nominal levels.

3. A permutation test statistic for general outcomes. In this section, we propose a response-adaptive design along with a large sample test statistic based on scores calculated from a general response variable. For each $j = 1, \ldots, n$, let r_{ij}, for $i \le j$, be the rank of the ith patient based on some outcome variable after j outcomes are available, where a larger rank indicates a better response to treatment. Define scores a_{ij} to be some function of the r_{ij}, $l \le i \le j \le n$, where

$\sum_{i=1}^{j} a_{ij} = 0, j = 1, \ldots, n$. Define $a_{ij}^{+} = a_{ij} I(a_{ij} > 0)$, where I is the indicator function, and as before, let $\Im_j = \sigma(Y_1, \ldots, Y_j)$, with Y_j defined as in Section 2. Let

$$\tilde{p}_i = E(Y_i \mid \Im_{i-1}) = \frac{1}{2}\left(1 + \frac{\sum_{j=1}^{i-1} a_{j,i-1}\left(Y_j - \frac{1}{2}\right)}{\sum_{j=1}^{i-1} a_{j,i-1}^{+}}\right),$$

$i = 1, 2, \ldots$. The better the responses of previous patients on treatment A are, relative to those on treatment B, the larger will be the value of p_i, the probability that patient i is assigned to A.

Proceeding as before, we define the array $\left\{\tilde{b}_{jn}\right\}$ to make

$$\sum_{j=1}^{n} a_{jn}\left(Y_j - \frac{1}{2}\right) = \sum_{j=1}^{n} \tilde{b}_{jn}(Y_j - \tilde{p}_j)$$

for each n. The test statistic

$$\tilde{T}_n = 2\,\frac{\sum_{j=1}^{n} a_{jn}\left(Y_j - \frac{1}{2}\right)}{\left(\sum_{j=1}^{n} \tilde{b}_{jn}^2\right)^{\frac{1}{2}}}$$

is then equivalent to

$$\tilde{W}_{nn} = 2\,\frac{\sum_{j=1}^{n} \tilde{b}_{jn}(Y_j - \tilde{p}_j)}{\left(\sum_{j=1}^{n} \tilde{b}_{jn}^2\right)^{\frac{1}{2}}}.$$

In the present case, however, the $\tilde{b}_{jn}$ depend not just on $\{a_{jn}\}$, but also on $\{a_{jk}\}$ for $k < n$. We have

(3.1)

$$\tilde{b}_{jn} = a_{jn} + \sum_{i=j+1}^{n} a_{jn} h_j(i),$$

$j = 1, \ldots, n$, where $h_j(i)$ is defined for $i \geq j$ by the recursion

(3.2)

$$\begin{aligned} h_i(i) &\equiv 1, \\ h_j(i) &= \sum_{k=j}^{i-1} \frac{a_{jk}}{2\sum_{l=1}^{k} a_{lk}^{+}} h_{k+1}(i), \quad i > j. \end{aligned}$$

We define

$$\tilde{S}_n \equiv \sum_{j=1}^{n} a_{jn} \left(Y_j - \frac{1}{2} \right).$$

The analog of conditions (2.2) and (2.3) are now given. Under the following two conditions,

(3.3)

$$\max_{l \leq j \leq n} \frac{\tilde{b}_{jn}^2}{\sum_{k=1}^{n} \tilde{b}_{kn}^2} \to 0 \text{ as } n \to \infty$$

and

(3.4)

$$\tilde{S}_n \,/\, \sum_{j=1}^{n} a_{jn}^{+} \to 1/2 \text{ in probability as } n \to \infty,$$

$\tilde{T}_n$ converges in distribution to a standard normal variate.

As before, we cannot guarantee that (3.3) and (3.4) will hold for every conceivable sequence $\{a_{ij}\}$. Thus, as in Section 2, we consider the case of responses generated by a probability mechanism to study the behavior of (3.3) and (3.4) under "average" conditions.

Assume now that the responses arise from an independent sequence $X_1, X_2, \ldots$ with a continuous distribution. Let R_{ij} be the rank of the ith patient after j responses are available.

EXAMPLE 3.1. *Simple rank scores.* For j even, these are defined by

$$A_{ij} = 8\left(\frac{R_{ij}}{j} - \frac{j+1}{2j}\right), \; i \leq j = 1, \; \ldots, \; n.$$

With this normalization, $\sum a_{ij}^{+} = j$. Let $\tilde{B}_{jn}$ and $H_j(i)$ denote the random analogs of $\tilde{b}_{jn}$ and $h_j(i)$ in (3.1) and (3.2), respectively.

THEOREM 3.1. *When the responses are generated from an independent continuous sequence,* $var\left(\tilde{S}_n/n\right) \to 0$ *as* $n \to \infty$, *where the expectation is taken over both the* $\{Y_j\}$ *and the scores.*

The proof of Theorem 3.1 can be found in Rosenberger (1992). Thus (3.4) holds in some average sense. Due to the complicated expression for the $\left\{\tilde{b}_{jn}\right\}$ in (3.1) and (3.2), we have not been able to show that (3.3) holds in the same sense; however, simulation evidence strongly suggests that (3.3) holds for this case.

Further simulations were designed to check condition (3.3) and the rate of convergence of $\tilde{T}_n$ to normality. The results are conservative in the extreme tails, especially for $n = 30$, but overall, a reasonable approximation to normality is demonstrated. Condition (3.3) appears to be holding at about the same rate $[O(n^{-1})]$ as $\max\left\{A_{jn}^2/\sum A_{jn}^2\right\}$.

EXAMPLE 3.2. *Van der Waerden scores.* The Van der Waerden scores are asymptotically equivalent to normal scores [see Lehmann (1975) p. 97], but are more tractable mathematically. They are defined by

$$A_{ij} = \Phi^{-1}\left(\frac{R_{ij}}{j+1}\right),$$

where Φ is the standard normal distribution function. For these scores, the behavior of $\max\left\{\tilde{B}_{jn}^2/\sum \tilde{B}_{kn}^2\right\}$ is even closer to that of

$$\max\left\{A_{jn}^2/\sum A_{jn}^2\right\},\ [O(n^{-1})],$$

than for simple ranks. It can be shown that Theorem 3.1 holds for Van der Waerden scores as well as simple rank scores [Rosenberger (1992)].

References

ATHREYA, K.B. and NEY, P.E. (1972). *Branching Processes.* New York: Springer-Verlag.

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

BEGG, C.B. (1990). On inferences from Wei's biased coin design for clinical trials. *Biometrika* **77** 467-484, with discussion.

CORNELL, R.G., LANDENBERGER, B.D., and BARTLETT, R.H. (1986). Randomized play the winner clinical trials. *Communication in Statistics - Theory and Methods* **15** 159-178.

FREEDMAN, D. (1965). Bernard Friedman's urn. *The Annals of Mathematical Statistics* **36** 956-970.

FRIEDMAN, B. (1949). A simple urn model. *Communication in Pure and Applied Mathematics* **2** 59-70.

HALL, P. and HEYDE, C.C. (1980). *Martingale Limit Theory and Its Application.* San Diego: Academic Press.

LEHMANN, E.L. (1975). *Nonparametrics: Statistical Methods Based on Ranks.* San Francisco: Holden-Day.

POCOCK, S.J. (1979). Allocation of patients to treatments in clinical trials. *Biometrics* **35** 183-197.

ROSENBERGER, W.F. (1992). *Asymptotic Inference Problems Arising From Clinical Trials Using Response-Adaptive Treatment Allocation*, Graduate School of Arts and Sciences, The George Washington University (doctoral dissertation).

ROSENBERGER, W.F. (1993). Asymptotic inference with response-adaptive treatment allocation designs. *The Annals of Statistics* **21** 2098-2107.

ROSENBERGER, W.F. and LACHIN, J.M. (1993). The use of response-adaptive designs in clinical trials. *Controlled Clinical Trials* **14** 471-484.

ROYALL, R.M. (1991). Ethics and statistics in randomized clinical trials (with discussion). *Statistical Science* **6** 52-62.

SIMON, R. (1972). Adaptive treatment assignment methods and clinical trials. *Biometrics* **33** 743-749.

SIMON, R. (1991). A decade of progress in statistical methodology for clinical trials. *Statistics in Medicine* **10** 1789-1817.

WEI, L.J. (1979). The generalized Pólya urn for sequential medical trials. *The Annals of Statistics* **7** 291-296.

WEI, L.J. (1988). Exact two-sample permutation tests based on the randomized play-the-winner rule. *Biometrika* **75** 603-606.

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.

WEI, L.J, SMYTHE, R.T., LIN, D.Y. and PARK, T.S. (1990). Statistical inference with data-dependent treatment allocation rules. *Journal of the American Statistical Association* **85** 156-162.

DEPARTMENT OF STATISTICS/STATISTICAL COMPUTING
THE BIOSTATISTICS CENTER
THE GEORGE WASHINGTON UNIVERSITY
6110 EXECUTIVE BOULEVARD, SUITE 750
ROCKVILLE, MD 20852