Sankhyā : The Indian Journal of Statistics
2001, Volume 63, Series B, Pt. 1, pp. 27–42

ROBUST ADAPTIVE DESIGNS IN CLINICAL TRAILS FOR CONTINUOUS RESPONSES

By ATANU BISWAS
and
AYANENDRANATH BASU
Indian Statistical Institute, Kolkata, India

SUMMARY. We consider the clinical trials scenario where patients enter the trial sequentially, and the experimenter has to adaptively select the better of the two competing treatments for future applications. This is particularly important since the subjects are human patients. However in almost all such techniques currently available, the responses are assumed to be dichotomous (success or failure). If the responses are continuous, one may consider dichotomizing them by setting an appropriate threshold value. But this may not have any real basis except its mathematical tractability. In this paper we introduce an adaptive design which uses all the information from the previous continuous responses for the next allocation without dichotomizing the responses. The method may be suitably modified to handle problems of lack of robustness (which may occur when the actual continuous measurements are used), as well as some other possible contingencies arising in such situations, such as delayed response.

1. Introduction

The problem of comparing two treatments A and B, say, is one of the most fundamental problems which statisticians have to deal with. It is a particularly important problem in the clinical trials scenario – for ethical reasons – where the experimenter has to select one of two competing treatments for future patients (human beings). These may be control and placebo or two competing treatments or therapies. Quite often one is the existing treatment and the other is the newly introduced one. As a real life example we consider the ACTG 076 trial conducted by the AIDS Clinical Trial Group (ACTG) (see Connor *et al.*, 1995, for details) to examine whether the drug zidovudine (AZT) can reduce vertical HIV transmission from the infected mothers to their infants. Out of 476 pregnant women who enrolled in the trial, a conventional stratified permuted block design assigned 238 women and their infants to the AZT group and the remaining 238 to the placebo group. Yao and Wei (1996) showed that instead of the permuted block design a suitable

Paper received November 1999; revised July 2000.
AMS (1991) *subject classification.* Primary 62L05; secondary 62F35.
Key words and phrases. Adaptive designs; clinical trials; delayed response; robustness.

randomized play-the-winner (RPW) rule, a particular adaptive design, could have done the allocation in a 300:176 ratio and 11 newborns could have been saved in the process of decision making without significant loss of efficiency. As the subjects are human beings, from an ethical point of view it is important that the rule be such that it allocates a fewer number of patients to the inferior treatment in course of decision making. For a sequential chain of patients entering a clinical trial, several data- dependent adaptive allocation rules are available for this purpose.

Modifying Zelen's (1969) pioneering concept of play-the-winner (PW) rule, Wei and Durham (1978) and Wei (1979) introduced randomized play-the-winner (RPW) rule. Other adaptive designs available in literature include the success driven design (SDD) (see Li, 1995; Durham, Flournoy and Li, 1998). Almost all such designs are essentially urn designs where responses are assumed to be dichotomous (success or failure). Responses obtained in the continuous scale have to be dichotomized by setting an appropriate threshold value leading to an obvious loss in information (see Bandyopadhyay and Biswas, 1996, 1997a, 1997b).

Some real life applications of adaptive designs are available in the literature. Iglewicz (1983) reports one use of data-dependent allocation in an unpublished application by M. Zelen to a lung cancer trial. The RPW rule, a particular adaptive design, has been used in at least three major clinical trials: the Michigan ECMO trial (see Bartlett *et al.*, 1985), and two trials of fluoxetine in depression to treat out-patients sponsored by Eli Lilly (see Tamura *et al.*, 1994). Ware (1989) and Rout *et al.* (1993) report on some other adaptive clinical trials. The number of real life adaptive designs, however, is not very large over the span of three decades. Main reasons for this may be (i) the failure of the ECMO trail through an 11:1 allocation, and (ii) lack of proper bridge between statisticians and clinical trial practitioners.

Faries *et al.* (1995) rightly feel that the dramatic advances in the computer technology and data access in the past decade makes the logistics of the adaptive trials much more feasible. This continuing advancement will force clinical trials to be more adaptive in all respects and statisticians should actively influence the design and analysis of such trials.

In the present paper we discuss an entirely new adaptive design (see Bandyopadhyay and Biswas, 2001) which uses all the information from the previous continuous responses for the next allocation. As already mentioned, there may not be any real basis for dichotomizing the responses, specially when well dispersed quantitative measurements are easily available. For example, in the fluoxetine trial (see Tamura *et al.*, 1994) the responses were reduction in $HAMD_{17}$ score which were measured in an ordinal scale. In survival analysis it could be more appropriate to use the total survival time instead of using an indicator setting a threshold. To our knowledge, adaptive allocation designs based on the actual continuous observations have not been attempted before except for a few attempts of discretization in nonparametric formulation using rank scores (Rosenberger, 1993; Bandyopadhyay and Biswas, 1999). Eisele (1994) used a doubly adaptive design for a biased coin type allocation of subjects to two treatments which took account of the current proportion of subjects assigned to each treatment and the current estimate of the desired allocation proportion obtained using all the continuous observations. Our present approach is

not of the biased coin type, and quite different from Eisele.

We introduce a robustness component to our design and decision making. This is extremely crucial when we use continuous responses for allocation as it will result in a big loss from an ethical point of view if a few observations (perhaps even a single one) lead to a large bunch of patients being assigned to a poor treatment when a better one is available. We use two types of robust estimators in our analysis, one being the M-estimator based on the well known Huber's ψ function; this we have used for the normal model. The other is a robust weighted likelihood estimator introduced by Field and Smith (1995) which we have used for the exponential model. Further we indicate possible extensions of our method to handle the delayed response case, which is inevitable, for example, in the fluoxetine trial discussed above.

In Section 2, we illustrate the continuous adaptive design and discuss its properties. Numerical computations are provided in Sections 3 which illustrates our present approach through simulations, as well as on a real life data set. Section 4 provides a discussion.

2. Sampling Scheme and Decision Rules

2.1 *Continuous adaptive design.* We denote the responses under treatments A and B by the real valued random variables X and Y, where we assume that the corresponding distribution functions F_1 and F_2 are continuous distribution functions (cdf's) possessing densities with respect to the Lebesgue measure, and that the expectations $\mu_i = \int x dF_i(x), i = 1, 2$, on which we will primarily focus our attention, exist. To make the problem and the analysis meaningful, we allocate the first patient to treatment A and the second one to treatment B. Starting from the third, incoming patients are allocated between treatments according to an adaptive allocation scheme which uses the information on the continuous responses of the previously allocated patients that are available till that point. Our allocation design will utilize the entire information (and not dichotomize it using some arbitrary cut-off point in the scale) and will be called the Continuous Adaptive Design (CAD).

We denote by X_i or Y_i the response for the i-th patient depending on whether it has been allocated to treatment A or B. Let δ_i be the indicator of assignment which takes the value 1 (0) if the i-th patient is allocated to treatment A (B) following the CAD design. (Notice that $\delta_1 = 1$ and $\delta_2 = 0$.) If we represent the entire sequence of observations by $\{W_i\}$, then $W_i = \delta_i X_i + (1 - \delta_i)Y_i$, $i = 1, \ldots, n$, where n is the maximum number of patients that may be treated. Our objectives in monitoring and analyzing the data thus generated are two fold: (i) to allocate the incoming patients adaptively – using all the information available till that point – to one of the two competing treatments such that on the long run more patients are assigned to the better treatment (this is important for ethical reasons); and (ii) to make inferences comparing the two populations, i.e. to declare that the sample evidence supports one of the three following hypotheses:

$$a_1 : \mu_1 = \mu_2, \quad a_2 : \mu_1 > \mu_2, \quad a_3 : \mu_1 < \mu_2. \tag{2.1}$$

The decision rule for this hypothesis is presented at the end of this section.

To explain the procedure of the analysis, we define, corresponding to the i-th incoming patient, the set of variables $\{\delta_i, W_i\}$, where W_i represents the actual continuous response of the i-th patient, and δ_i is the indicator variable defined earlier. The number of patients N_{Ak} and N_{Bk} treated by treatments A and B respectively just after the entry of the k-th patient are then given by

$$N_{Ak} = \sum_{i=1}^{k} \delta_i, \quad N_{Bk} = \sum_{i=1}^{k} (1 - \delta_i) = k - N_{Ak}. \tag{2.2}$$

Depending on the specific objectives of the experimenter and the nature of the problem, the final analysis may be performed after all n patients have been observed, or a provision may be retained for continuous monitoring and analysis with the possibility of early stopping.

We now describe our basic decision rule. When it is necessary to consider additional complications such as problems with outliers (Section 2.2), or, say delayed response, these have to be appropriately incorporated in the decision rule. The rule for allocating the $(k+1)$-th patient, having observed the responses $W_1, \ldots, W_k$ (if available) and the previous allocations is as follows:

(i) Choose a continuous cumulative distribution function $G(\cdot)$ which is symmetric about 0, i.e. $G(0) = 1/2$, $G(-x) = 1 - G(x)$ (such as the $N(0,1)$ cdf $\Phi(x)$).

(ii) Determine the observed sample means for the two populations at this stage as

$$\hat{\mu}_{Ak} = \frac{\sum_{i=1}^{k} \delta_i W_i}{\sum_{i=1}^{k} \delta_i}, \quad \hat{\mu}_{Bk} = \frac{\sum_{i=1}^{k} (1 - \delta_i) W_i}{\sum_{i=1}^{k} (1 - \delta_i)}. \tag{2.3}$$

(iii) Allocate the $(k+1)$th patient to treatment A with probability $G((\hat{\mu}_{Ak} - \hat{\mu}_{Bk})/c)$, and to treatment B with probability $(1 - G((\hat{\mu}_{Ak} - \hat{\mu}_{Bk})/c)$, where c is some appropriate scaling constant. The allocation procedure should favour the treatment which has led to larger responses on the average in the past – which this allocation rule manages to achieve. (See Bandyopadhyay and Biswas (2000, 2001) for such designs in some other situations like the testing of normal mean and the presence of covariates, respectively).

An additional refinement that may be possible and sometimes worthwhile is to choose the cdf G as a function of k, where k is the number of patients observed previously. The advantage of allowing this is that since big differences in $\hat{\mu}_{Ak} - \hat{\mu}_{Bk}$ are more meaningful for large k, one can choose G_k so that for any positive number $a > 0$, $G_k(a)$ is a steadily increasing function of k. Thus the same observed difference is taken as stronger evidence of the difference of means when k is larger.

Unlike Eisele (1994), no specific target imbalance is set for our design. This is in the same spirit of the RPW rule. In fact, as in the RPW rule, here also we can expect some limiting proportion of allocation. Intuitively, we expect the limiting proportion of allocation to A to be $G((\mu_A - \mu_B)/c)$. This has been proved to be true, in some special cases, by Bandyopadhyay and Biswas (2001), and currently more elaborate research is going on.

Under this scheme, therefore,

$$P(\delta_{k+1} = 1) = E(G(\hat{\mu}_{Ak} - \hat{\mu}_{Bk})/c) = \int G(x) dH_k^c(x), \tag{2.4}$$

where $\hat{\mu}_{Ak}$ and $\hat{\mu}_{Bk}$ are based on the first k patients, and $H_k^c(\cdot)$ is the corresponding distribution of $(\hat{\mu}_{Ak} - \hat{\mu}_{Bk})/c$ based on k patients.

Finally we carry out the decision making problem addressed in (2.1). Based on targeted n sample observations one chooses a_2 or a_3 if the observed $\hat{\mu}_{An} - \hat{\mu}_{Bn}$ is greater than a pre-specified cut-off point u, or less than $-u$ respectively. Otherwise one chooses a_1. One can carry out early stopping as well, as in most of the sequential procedures. In that case one has to choose any of the three actions as soon as the accumulated information is sufficient to do so. However, in the present article, our main focus is on illustrating the possible applicability of CAD with robust estimates, and we will not discuss the details of the early stopping in the decision making procedure; see Bandyopadhyay and Biswas (1996).

2.2 *Robustness considerations.* In the above scheme of allocation we take the actual continuous response of the patients into consideration. In doing that, we have the opportunity of possibly identifying outlying observations which can exert undue influence on our analysis. Under traditional designs such as the RPW, where the scheme allows only one or zero as a response it is impossible to identify an observation as an outlier by looking at the corresponding indicator variable.

But for the continuous adaptive design scheme, robustness considerations are appropriate and necessary. This is particularly so because the estimates in equation (2.3) estimating the unknown true means in question are the sample means of the responses observed till that point. Given the nonrobustness of the sample mean as the estimate of the population mean (although it is asymptotically efficient under many common parametric models) we will like to use robust analogs of the estimators in (2.3).

In this paper, we have used two different parametric robust estimates, one based on M-estimators with Huber's ψ function (see, for example, Hampel et al. 1986), which we have used for estimating the parameter in the normal location problem. In the one population case one solves for μ from the estimating equation $\sum_i \psi_b((X_i - \mu)/\hat{\sigma}) = 0$, where $X_1, \ldots, X_n$ represent the sampled observations, and $\hat{\sigma}$ is a robust estimate of the scale. The function $\psi_b(\cdot)$ has the form

$$\psi_b(x) = \begin{cases} x & \text{if } |x| \le b \\ b & \text{if } x > b \\ -b & \text{if } x < -b \end{cases} \tag{2.5}$$

so that this limits the impact of observations which have unusually large residuals. The other, used in the context of exponential models, is based on the robust weighted likelihood estimators of Field and Smith (1995) which estimates the parameter by iteratively solving a weighted likelihood equation for the parameter θ where the weights are given by

$$w(x, \theta) = \begin{cases} F(x, \theta)/p & \text{if } F(x, \theta) < p \\ 1 & \text{if } p \le F(x, \theta) \le 1 - p \\ (1 - F(x, \theta))/p & \text{if } F(x, \theta) > 1 - p \end{cases} \tag{2.6}$$

$F(x,\cdot)$ being the distribution function of the response under the model, θ is the current estimate of the parameter, and p is some suitable small value, $p \in (0,1)$.

Obviously, these are just a few of the possible robust methods that may be used in practice. These have been chosen to demonstrate the use of robust techniques. The user can, of course, select other appropriate methods of determining the mean depending on the problem at hand.

3. Numerical Studies

Unlike the case of the ordinary adaptive designs of the success failure type, theoretical calculations in case of a continuous adaptive design are far more complicated, and in this section we try to illustrate the properties and the performance of the method through extensive numerical studies including simulations and examples.

3.1 *Simulation results.* In our first study we use the exponential model. The responses in the two groups are assumed to be exponential with means λ_A and λ_B (denoted here by $exp(\lambda_A)$ and $exp(\lambda_B)$), respectively, $\lambda_A > \lambda_B$. Since our allocation scheme deterministically allocates the first patient to group A and the second patient to group B, we initially draw a random number from $exp(\lambda_A)$ and $exp(\lambda_B)$ each to denote these responses. For each of the successive allocations we use the current estimates $\hat{\lambda}_{Ak}$ and $\hat{\lambda}_{Bk}$ and determine the next allocation according to probabilities $G((\hat{\lambda}_{Ak} - \hat{\lambda}_{Bk})/c)$ for group A and $1 - G((\hat{\lambda}_{Ak} - \hat{\lambda}_{Bk})/c)$ for B where $G(\cdot) = \Phi(\cdot)$ is the standard normal distribution function. This is done for a fixed number n of patients and two different types of estimates for the λ parameters – the ordinary sample mean and the robust weighted likelihood estimate for λ based on the Field and Smith approach. For the Field and Smith approach we use $p = 0.01, 0.05$ and 0.1. For each approach we use several values of the scaling constant c. At the end of the trial, after all n patients have been thus allocated, we calculate, for each method, the final estimates of λ_A and λ_B. In addition, we calculate the total number of allocations T_A and T_B to the drugs A and B respectively ($T_A + T_B = n$). We also calculate the loss corresponding to the final decision with an appropriate loss function. This loss function is chosen as follows: We take the loss for choosing a_i when a_i is true as 0. Loss for choosing any decision one step apart (choosing a_2 or a_3 when a_1 is true; choosing a_1 when a_2 or a_3 is true) is 1, while loss for taking a decision two step apart (choosing a_2 when a_3 is true or vice versa) is L (≥ 1). We refer to Ferguson (1967) in connection with such a loss function.

We repeat this process over a large number of samples, and calculate the means and variances of the number of allocations to either of the two treatments. In addition we calculate the sample estimates of the probabilities of actions a_3 and a_1, denoted, respectively as Prob1 and Prob2, and calculate the observed risk over the different samples. Different choices for the λ values and different values of n are used to gain a better understanding of the performance of the method. We present some of these calculations in Tables 1-2. The patterns of the results of other calculations are similar. In these two tables, as well as the later ones presented here, n equals 20, and the number of replications equals 200.

Table 1. Simulation results for the exponential model. Data for populations A, B are generated from $exp(\lambda_A)$ and $exp(\lambda_B)$ distributions. T_A is the number of patients (out of 20) allocated to population A. Number of replications is 200. The estimator used is the sample mean. Prob1, Prob2 are the empirical probabilities of the decisions $a_3 : \mu_1 < \mu_2$, $a_1 : \mu_1 = \mu_2$ for the given cut-offs.

				cut-off=1			cut-off=2		
(λ_A, λ_B)	c	$E(T_A)$	$V(T_A)$	Prob1	Prob2	Risk at $L=1$	Prob1	Prob2	Risk at $L=1$
	5	9.785	4.8832	0.005	0.985	0.015	0.000	1.000	0.000
(1.0,1.0)	10	10.035	4.8882	0.005	0.960	0.040	0.000	1.000	0.000
	20	10.025	4.3964	0.015	0.970	0.030	0.000	1.000	0.000
	5	9.605	5.0643	0.055	0.940	0.940	0.000	1.000	1.000
(1.0,1.2)	10	9.730	4.8212	0.055	0.940	0.945	0.000	1.000	1.000
	20	9.790	3.9155	0.050	0.945	0.950	0.000	1.000	1.000
	5	8.255	6.5125	0.515	0.485	0.485	0.075	0.925	0.925
(1.0,2.0)	10	9.145	5.3507	0.505	0.495	0.495	0.045	0.955	0.955
	20	9.485	5.2259	0.465	0.535	0.535	0.075	0.925	0.925
	5	6.095	6.2573	0.965	0.035	0.035	0.790	0.210	0.210
(1.0,4.0)	10	7.980	5.4870	0.960	0.040	0.040	0.755	0.245	0.245
	20	8.615	4.8008	0.945	0.055	0.055	0.795	0.205	0.205

Table 2. Simulation results for the exponential model. Data for populations A and B are generated from $exp(\lambda_A)$ and $exp(\lambda_B)$ distributions. T_A is the number of patients (out of 20) allocated to population A. Number of replications is 200. The estimator used is the weighted likelihood estimator ($p = 0.05$). Prob1, Prob2 are the empirical probabilities of the decisions $a_3 : \mu_1 < \mu_2$, $a_1 : \mu_1 = \mu_2$ for the given cut-offs.

				cut-off=1			cut-off=2		
(λ_A, λ_B)	c	$E(T_A)$	$V(T_A)$	Prob1	Prob2	Risk at $L=1$	Prob1	Prob2	Risk at $L=1$
	5	10.035	4.5666	0.005	0.975	0.025	0.000	1.000	0.000
(1.0,1.0)	10	9.800	4.1508	0.020	0.950	0.050	0.000	1.000	0.000
	20	10.025	4.3361	0.030	0.940	0.060	0.000	1.000	0.000
	5	9.740	4.8366	0.055	0.920	0.945	0.005	0.995	0.995
(1.0,1.2)	10	9.815	4.1917	0.080	0.900	0.920	0.005	0.995	0.995
	20	9.830	3.9207	0.075	0.890	0.925	0.000	1.000	1.000
	5	8.615	6.3686	0.455	0.535	0.545	0.070	0.930	0.930
(1.0,2.0)	10	9.215	4.8329	0.465	0.535	0.535	0.045	0.955	0.955
	20	9.525	4.1803	0.435	0.550	0.565	0.095	0.905	0.905
	5	6.440	6.4386	0.920	0.080	0.080	0.715	0.285	0.285
(1.0,4.0)	10	8.255	5.1960	0.885	0.115	0.115	0.680	0.320	0.320
	20	9.040	4.0285	0.880	0.120	0.120	0.675	0.325	0.325

The same procedure is then repeated, but now the response from population A is assumed to have a distribution which is a mixture of exponentials of the form $0.9exp(\lambda_A) + 0.1exp(\lambda_C)$, for a suitably chosen value of λ_C. The second component in this population, however, is considered to be a contaminant, and in this case we are interested in observing whether the procedures can detect that population B is better than the major component of population A, or whether the presence of the second component causes a substantial increase in the proportion of allocations to drug A (which we do not want to happen). We expect that in this respect the robust procedures will be fairly stable. As before, we present some of the computations keeping parity with Tables 1-2. These are presented in Tables 3-4 with $\lambda_C = 2$.

Table 3. Simulation results for the exponential model. Data for populations A and B are generated from $0.9exp(\lambda_A) + 0.1exp(2)$ and $exp(\lambda_B)$ distributions. T_A is the number of patients (out of 20) allocated to treatment A. Number of replications is 200. The estimator used is the population mean. Prob1, Prob2 are the empirical probabilities of the decisions $a_3 : \mu_1 < \mu_2$, $a_1 : \mu_1 = \mu_2$ for the given cut-offs.

				cut-off=1			cut-off=2		
(λ_A, λ_B)	c	$E(T_A)$	$V(T_A)$	Prob1	Prob2	Risk at $L=1$	Prob1	Prob2	Risk at $L=1$
	5	9.950	6.1080	0.020	0.970	0.030	0.000	1.000	0.000
(1.0,1.0)	10	10.020	5.3564	0.005	0.970	0.030	0.000	1.000	0.000
	20	9.880	4.2669	0.020	0.945	0.055	0.000	1.000	0.000
	5	9.815	5.8098	0.050	0.925	0.950	0.000	1.000	1.000
(1.0,1.2)	10	9.855	4.9789	0.025	0.960	0.975	0.000	1.000	1.000
	20	9.870	4.7870	0.040	0.940	0.960	0.000	1.000	1.000
	5	8.555	6.5598	0.425	0.575	0.575	0.090	0.910	0.910
(1.0,2.0)	10	9.155	5.2070	0.430	0.560	0.570	0.065	0.935	0.935
	20	9.600	4.5126	0.470	0.530	0.530	0.085	0.915	0.915
	5	6.620	7.1815	0.940	0.040	0.060	0.790	0.210	0.210
(1.0,4.0)	10	8.130	6.1137	0.890	0.110	0.110	0.735	0.265	0.265
	20	8.770	4.4091	0.935	0.065	0.065	0.750	0.250	0.250

Table 4. Simulation results for the exponential model. Data for populations A, B are generated from $0.9exp(\lambda_A) + 0.1exp(2)$ and $exp(\lambda_B)$ distributions. T_A is the number of patients (out of 20) allocated to treatment A. Number of replications is 200. The estimator used is the weighted likelihood estimator with $p = 0.05$. Prob1, Prob2 are the empirical probabilities of the decisions $a_3 : \mu_1 < \mu_2$, $a_1 : \mu_1 = \mu_2$ for the given cut-offs.

				cut-off=1			cut-off=2		
(λ_A, λ_B)	c	$E(T_A)$	$V(T_A)$	Prob1	Prob2	Risk at $L=1$	Prob1	Prob2	Risk at $L=1$
	5	9.940	5.4235	0.045	0.935	0.065	0.000	1.000	0.000
(1.0,1.0)	10	9.920	5.2398	0.030	0.940	0.060	0.000	1.000	0.000
	20	9.880	4.2971	0.030	0.920	0.080	0.000	1.000	0.000
	5	9.880	6.3172	0.050	0.915	0.950	0.000	1.000	1.000
(1.0,1.2)	10	9.820	4.7815	0.075	0.895	0.925	0.005	0.995	0.995
	20	9.870	4.9277	0.065	0.910	0.935	0.000	1.000	1.000
	5	8.525	5.8486	0.455	0.530	0.545	0.085	0.910	0.915
(1.0,2.0)	10	9.165	4.6209	0.460	0.535	0.540	0.065	0.935	0.935
	20	9.630	4.2443	0.400	0.595	0.600	0.055	0.945	0.945
	5	6.610	7.0532	0.885	0.105	0.115	0.635	0.365	0.365
(1.0,4.0)	10	7.925	5.8285	0.870	0.125	0.130	0.635	0.365	0.365
	20	8.140	11.2300	0.780	0.220	0.220	0.605	0.395	0.395

This entire process is then replicated for the normal model, where the responses are assumed to have $N(\mu_A, \sigma^2)$ and $N(\mu_B, \sigma^2)$ distributions, $\mu_A \leq \mu_B$. The method assumes the variances to be equal but unknown. (For generating the simulated data we have taken $\sigma^2 = 1$.) In this case the allocation probabilities are $G((\hat{\mu}_{Ak} - \hat{\mu}_{Bk})/c)$ and $1 - G((\hat{\mu}_{Ak} - \hat{\mu}_{Bk})/c)$ for populations A and B respectively. For this problem we calculate the estimates of μ_A and μ_B through the sample means, and the M-estimates using the Huber ψ function with several values of b. At the entry of the k-the patient, the robust estimate of scale $\hat{\sigma}$ is chosen to be

$$\hat{\sigma} = median\{|U_i - median(U_i)|, i = 1, \ldots, k_1, |V_j - median(V_j)|, j = 1, \ldots, k_2\}/0.674,$$

where k_1 and k_2 are the number of patients allocated to A and B, $k_1 + k_2 = k$, and the corresponding observations in groups A and B are $U_1, \ldots, U_{k_1}$, and $V_1, \ldots, V_{k_2}$ respectively. However if either k_1 or k_2 is 1, the estimate $\hat{\sigma}$ is obtained only from the other group. Some of the results with sample means and M-estimates (with tuning parameter $b = 1.5$) are presented in Tables 5-6. This process is then repeated where the response from population A has the distribution $0.9N(\mu_A, 1) + 0.1N(\mu_C, 1)$, the second component representing the contamination. The results are given in Tables 7-8 with $\mu_C = 10$.

Table 5. Simulation results for the normal model. Data for populations A, B are generated from $N(\mu_A, 1)$ and $N(\mu_B, 1)$ distributions. T_A is the number of patients (out of 20) allocated to treatment A. Number of replications is 200. The estimator used is the sample mean. Prob1, Prob2 are the empirical probabilities of the decisions $a_3 : \mu_1 < \mu_2$, $a_1 : \mu_1 = \mu_2$ for the given cut-offs.

				cut-off=0.5			cut-off=2		
(μ_A, μ_B)	c	$E(T_A)$	$V(T_A)$	Prob1	Prob2	Risk at $L = 1$	Prob1	Prob2	Risk at $L = 1$
	5	10.010	5.4672	0.185	0.685	0.315	0.000	1.000	0.000
(1.0,1.0)	10	10.045	4.0030	0.185	0.705	0.295	0.000	1.000	0.000
	20	10.030	3.9790	0.175	0.705	0.295	0.000	1.000	0.000
	5	9.765	4.4922	0.265	0.685	0.735	0.000	1.000	1.000
(1.0,1.2)	10	9.815	4.4229	0.265	0.665	0.735	0.000	1.000	1.000
	20	9.985	3.9143	0.260	0.685	0.740	0.000	1.000	1.000
	5	8.390	4.8220	0.900	0.100	0.100	0.010	0.990	0.990
(1.0,2.0)	10	9.215	4.6319	0.910	0.090	0.090	0.000	1.000	1.000
	20	9.825	3.9139	0.860	0.140	0.140	0.025	0.975	0.975
	5	5.765	3.3365	1.000	0.000	0.000	0.990	0.010	0.010
(1.0,4.0)	10	7.815	4.7646	1.000	0.000	0.000	0.985	0.015	0.015
	20	8.985	5.0098	1.000	0.000	0.000	0.985	0.015	0.015

Table 6. Simulation results for the normal model. Data for populations A, B are generated from $N(\mu_A, 1)$ and $N(\mu_B, 1)$ distributions. T_A is the number of patients (out of 20) allocated to treatment A. Number of replications is 200. The estimator used is the Huber's M-estimator. Prob1, Prob2 are the empirical probabilities of the decisions $a_3 : \mu_1 < \mu_2$, $a_1 : \mu_1 = \mu_2$ for the given cut-offs.

				cut-off=0.5			cut-off=2		
(μ_A, μ_B)	c	$E(T_A)$	$V(T_A)$	Prob1	Prob2	Risk at $L = 1$	Prob1	Prob2	Risk at $L = 1$
	5	10.190	5.6220	0.165	0.700	0.300	0.000	1.000	0.000
(1.0,1.0)	10	10.025	4.0546	0.190	0.675	0.325	0.000	1.000	0.000
	20	10.045	4.1236	0.190	0.680	0.320	0.000	1.000	0.000
	5	9.835	4.3797	0.275	0.650	0.725	0.000	1.000	1.000
(1.0,1.2)	10	9.810	4.4964	0.280	0.665	0.720	0.000	1.000	1.000
	20	10.025	4.0446	0.285	0.650	0.715	0.000	1.000	1.000
	5	8.260	4.3240	0.890	0.110	0.110	0.030	0.970	0.970
(1.0,2.0)	10	9.185	4.5636	0.895	0.105	0.105	0.005	0.995	0.995
	20	9.875	3.5773	0.850	0.150	0.150	0.010	0.990	0.990
	5	5.800	3.9397	1.000	0.000	0.000	0.985	0.015	0.015
(1.0,4.0)	10	7.845	4.9558	1.000	0.000	0.000	0.985	0.015	0.015
	20	8.890	4.7617	1.000	0.000	0.000	0.995	0.005	0.005

Table 7. Simulation results for the normal model. Data for populations A, B are generated from $0.9N(\mu_A, 1) + 0.1N(10, 1)$ and $N(\mu_B, 1)$ distributions. T_A is the number of patients (out of 20) allocated to treatment A. Number of replications is 200. The estimator used is the sample mean. Prob1, Prob2 are the empirical probabilities of the decisions $a_3 : \mu_1 < \mu_2$, $a_1 : \mu_1 = \mu_2$ for the given cut-offs.

				cut-off=0.5			cut-off=2		
(μ_A, μ_B)	c	$E(T_A)$	$V(T_A)$	Prob1	Prob2	Risk at $L = 1$	Prob1	Prob2	Risk at $L = 1$
	5	10.635	6.7053	0.260	0.415	0.585	0.000	0.890	0.110
(1.0,1.0)	10	10.575	4.9391	0.290	0.380	0.620	0.000	0.910	0.090
	20	10.295	4.4100	0.350	0.350	0.650	0.000	0.920	0.080
	5	10.345	5.9960	0.145	0.380	0.855	0.000	0.955	1.000
(1.0,1.2)	10	10.195	5.2331	0.160	0.395	0.840	0.000	0.920	1.000
	20	10.160	4.5974	0.105	0.440	0.895	0.000	0.945	1.000
	5	8.840	6.6376	0.495	0.335	0.505	0.020	0.975	0.980
(1.0,2.0)	10	9.840	4.9049	0.410	0.400	0.590	0.005	0.980	0.995
	20	9.875	3.9994	0.490	0.380	0.510	0.020	0.970	0.980
	5	6.390	6.5004	0.930	0.070	0.070	0.685	0.315	0.315
(1.0,4.0)	10	8.085	4.9123	0.935	0.055	0.065	0.630	0.370	0.370
	20	8.860	4.5230	0.955	0.040	0.045	0.695	0.305	0.305

Table 8. Simulation results for the normal model. Data for populations A, B are generated from $0.9N(\mu_A, 1) + 0.1N(10, 1)$ and $N(\mu_B, 1)$ distributions. T_A is the number of patients (out of 20) allocated to treatment A. Number of replications is 200. The estimator used is the Huber M-estimator ($b = 1.5$). Prob1, Prob2 are the empirical probabilities of the decisions $a_3 : \mu_1 < \mu_2$, $a_1 : \mu_1 = \mu_2$ for the given cut-offs.

				cut-off=0.5			cut-off=2		
(μ_A, μ_B)	c	$E(T_A)$	$V(T_A)$	Prob1	Prob2	Risk at $L = 1$	Prob1	Prob2	Risk at $L = 1$
	5	10.430	5.6433	0.195	0.650	0.350	0.000	1.000	0.000
(1.0,1.0)	10	10.325	4.1803	0.180	0.640	0.360	0.000	0.995	0.005
	20	10.040	4.1290	0.185	0.710	0.290	0.000	1.000	0.000
	5	10.020	5.3966	0.205	0.615	0.795	0.000	1.000	1.000
(1.0,1.2)	10	10.040	4.9029	0.195	0.715	0.805	0.000	1.000	1.000
	20	9.955	4.1537	0.130	0.750	0.870	0.000	1.000	1.000
	5	8.585	5.0882	0.765	0.220	0.235	0.005	0.990	0.995
(1.0,2.0)	10	9.260	4.1833	0.765	0.220	0.235	0.015	0.985	0.985
	20	9.675	4.5119	0.750	0.245	0.250	0.010	0.990	0.990
	5	6.310	4.5366	1.000	0.000	0.000	0.905	0.095	0.095
(1.0,4.0)	10	7.875	4.3511	1.000	0.000	0.000	0.930	0.070	0.070
	20	8.640	4.0506	1.000	0.000	0.000	0.950	0.050	0.050

From the computations we observe that the adaptive design we have used is doing satisfactorily as the main objective of any adaptive allocation design is fulfilled here, i.e., on an average a larger number of experimental units are being treated by the eventual winner when the data are without contamination. Moreover it uses the continuous responses with due weightage.

From Tables 7-8 we find that the robust method performs better in allocating to the more effective treatment when contamination is present in the data. The sample mean consistently allocates more patients to the inferior treatment compared to the robust estimate (Tables 7 and 8). Notice that in this case the contaminating component has a much larger mean. However the robust estimator performs only

marginally better (if at all) than the sample mean in Tables 3 and 4, since the contaminating component increases the mean of the first population only slightly.

To illustrate the applications of the method for larger sample sizes, Tables 1-4 were reconstructed for samples of size 50. The results were similar to the case $n = 20$. For brevity the entire sets of numbers are not reproduced here. However for each of the four different situations corresponding to Tables 1–4, with the set of parameters being $\lambda_A = 1$, $\lambda_B = 2$, $\lambda_c = 2$, and $c = 5$, the histogram of the entire set of 200 different allocations to treatment A are shown in Figure 1. There are no cases with such extremely small allocations to one arm as in the controversial ECMO trial.

Thus the robust methods can perform better than the sample mean in terms of satisfactory allocation to the superior treatment. The risk decreases when the difference between the treatments increases in the sense that the proposed procedure catches the difference with higher probability. Again any adaptive procedure must have some variability in allocation, although the expected allocation is lager to the better treatment. It is also observed that the variance observed in the present case is not too much in comparison to the standard RPW or PW rule. In none of the simulated cycles we got a pattern where only one observation was assigned to one of the treatments, such as the controversial ECMO trial.

3.2 *An example.* As CAD is a newly introduced sampling design, it is not possible to have a real life dataset based on CAD in the existing literature. Thus to have a meaningful comparison we consider a part of the dataset obtained from the fluoxetine trial, a famous real life adaptive trial available in literature. We consider the responses by the two treatments obtained there in two stacks, and carry out our CAD based on those responses. A particular permutation of the first 20 observations from each treatment is given as:

A : $4, 2, -20, 0, -21, -3, -16, -9, 3, 0, -6, -7, -3, -3, -4, -16, -6, -11, -3, -16,$

B : $-1, -1, -12, -2, -11, -17, -5, -12, -10, -21, -7,$
$-8, -20, -4, 2, -14, -1, -8, -16, -15.$

For our experiment when a response is required from a treatment we consider the stack for that treatment and take the response from the top of the remaining observations. As an illustration we consider the normal model with equal variance and estimate the means using the M-estimator with different values of the tuning parameter b. The results (number of allocations to treatment A) are presented in Table 9 along with the allocation obtained by using the sample mean. The experiment was terminated when 20 individuals had entered the study. Notice that the robust methods generally allocated more observations to treatment A, the treatment with the larger mean, compared to the method based on sample mean, probably because of the latter's inability to deal with the three very small values $-20, -21$ and -16, early on in the chain of treatment A values.

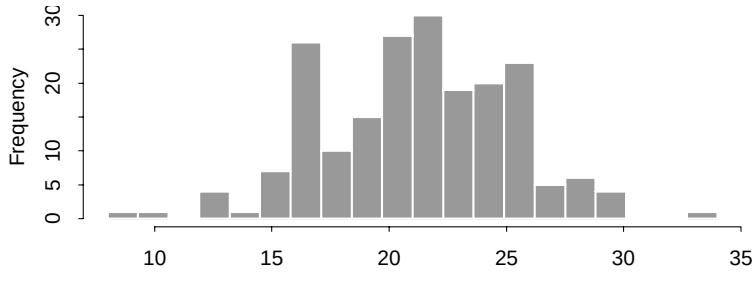

Figure 1a: The sample mean is used on uncontaminated data

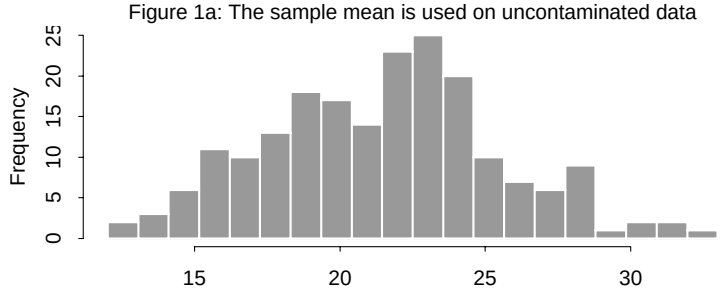

Figure 1b: The robust estimator is used on uncontaminated data

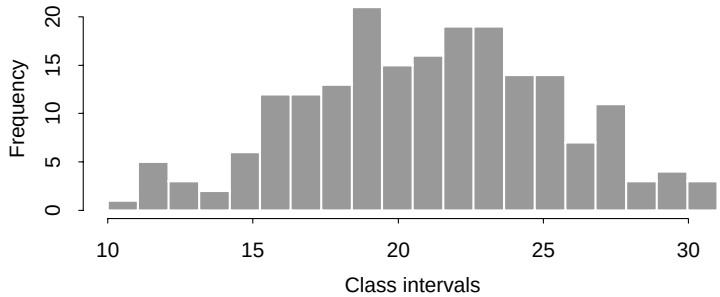

Figure 1c: The sample mean is used on contaminated data

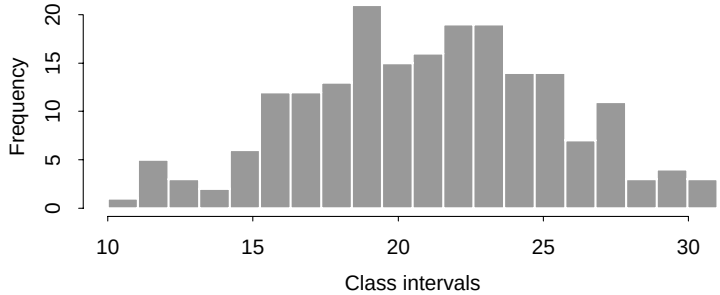

Figure 1d: The robust estimator is used on contaminated data

Figure 1: Histograms of the number of individuals (out of 50) allocated to the first treatment in 200 independent runs for four different situations.

Table 9. An example based on the permutations of the observations of the two treatment groups for the fluoxetine trial. Computations of the example are done assuming normal model. The entries represent the number of allocations to treatment A when different estimates for the parameters (M-estimates with different tuning parameters and the sample mean) are used.

Scaling constant c	Estimates			
	M-estimates with tuning parameter b			Sample mean
	$b = 1.25$	$b = 1.50$	$b = 2.00$	
2.5	12	10	10	10
5.0	11	11	11	10
7.5	11	11	10	10
10.0	11	10	10	9

4. **Discussions**

Although it is easy to handle the situation of instantaneous responses, often, in practice, we may face the situation of delayed response. In fact the fluoxetine trial by Tamura *et al.* (1994) is a real life example of a trial with delayed response. Delayed response is particularly important in adaptive trials where the existing data decides the allocation pattern of the next incoming patient. Some work involving delayed responses are available in the RPW rule. A model is provided by Wei (1988). Bandyopadhyay and Biswas (1996) have also incorporated such a possible delay in response in the RPW case. Then Biswas (1999) has revisited the problem. In such a case one can carry out our allocation procedure by obtaining $\hat{\mu}_{Ak}$ and $\hat{\mu}_{Bk}$ only on the basis of the available responses before the entry of the $(k+1)$st patient.

Alternatively we can use a suitable surrogate declaration variable provided its response is quick enough and the surrogate exhibits a high positive correlation with the true responses. Tamura *et al.* (1994) have used such a surrogate in their approach. But, in our case, it will increase theoretical complication as the joint distribution of the true response and the surrogate will now have to be considered.

Choice of the scaling constant c is of course an important question which is to be addressed with great caution. It controls the allocation pattern of incoming patients and distribute them according to the state of the art provided by the data at any stage. A small value of c may make the design more sensitive to outliers particularly during the earlier stages when data are not adequate. Large values of c will have a tendency to ignore the information provided by the data and will eventually pull the allocation ratio towards the 50:50 pattern. Figure 2 shows the allocation probabilities (at any stage) to the better treatment plotted against difference of the estimates of treatment means (standard or robust estimates) and the value of the scaling constant. It seems natural to demand that the choice of c should be linked to the standard deviation of the estimated mean difference: for example, if c is chosen to be a constant times the estimated standard deviation of the estimated mean difference, the allocation probabilities will be based on the scaled differences

of the estimated means. This will make the design more complex in the sense

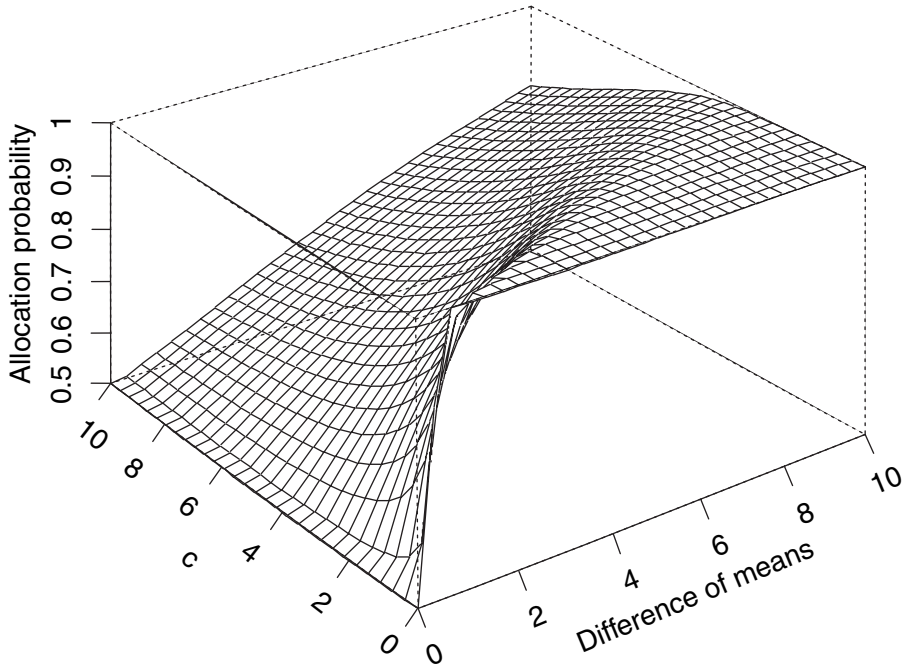

Figure 2: Probability of allocation to the better treatment against different estimated difference of means and c

that the value of c will change at the time of each allocation. Sometimes the experimenter may have prior idea about the degree of variability among the estimated mean difference (as functions of the sample sizes) which could also be used. For example when the treatment responses are assumed to be normal, often the variances are assumed to be known and equal. Also sometimes variances are functions of treatment means only, as in the exponential case. In general, one can talk about a sequence $\{c_n,\ n \geq 1\}$ instead of a fixed c which can be preset or can be adaptively obtained reflecting the current estimates of the variability from the data obtained till that point. The performance of the method under different possible choices of c is currently under investigation, and we hope to report this in a sequel paper.

The proposed CAD allocates a larger number of patients to the better treatment arm. Obviously, this complicated dependent procedure will reduce the power of the method compared to a fixed sample size design. Estimating the actual powers of this method in different situations will require large scale simulations which we hope to conduct in the future, but we have no reason to believe that the power of the method will be lower than that achieved through a RPW design.

In the present article we have taken the standard normal distribution function Φ as a choice of G. The familiarity of this distribution, coupled with the fact that

the quantiles of this distribution are now inbuilt in most statistical softwares make it a natural choice. In general however G needs to satisfy only the two properties that it is continuous and symmetric around zero. The allocation probabilities will be different for different choices of G, but by choosing c depending on G, this differences can be minimized. In general, the choice of c will be the more difficult problem than the choice of G.

Acknowledgment. The authors wish to thank the referees for their valuable suggestions which led to some improvement over an earlier version of the manuscript.

References

Bandyopadhyay, U. and Biswas, A. (1996). Delayed response in randomized play-the-winner rule: a decision theoretic outlook, *Calcutta Statistical Association Bulletin*, **46**, 69–88.

– – –– (1997a). Sequential comparison of two treatments in clinical trials: a decision theoretic approach based on randomized play-the-winner rule, *Sequential Analysis*, **16**, 65–91.

– – –– (1997b). Some sequential tests in clinical trials based on randomized play-the-winner rule. *Calcutta Statistical Association Bulletin*, **47**, 67–89.

– – – – (1999). Nonparametric two sample problem based on adaptive allocation using Wilcoxon-Mann-Whitney scores. *Technical Report No.* **ASD/99/9**, Indian Statistical Institute, Calcutta.

– – –– (2000). An alternative one sided test of normal mean, *Sequential Analysis*, **19**, 205-219.

– – – – (2001). Adaptive designs for normal responses with prognostic factors, *Biometrika*. To appear.

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 trial, *Pediatrics*, **76**, 479–487.

Biswas, A. (1999). Delayed response in randomized play-the-winner rule revisited, *Communications in Statistics, Simulation and Computation*, Vol. **28**(3), 715–731.

Connor, E.M., Sperling, R.S., Gelber, R., Kiselev, P., Scott, G., O'Sullivan, M.J., Vandyke, R., Bey, M., Shearer, W., Jacobson, R.L., Jiminez, E., O'Neill, E., Bazin, B., Delfraissy, J., Culname, M., Coombs, R., Elkins, M., Moye, J., Stratton, P. and Balsey, J. (1994). Reduction of maternal-infant transmission of human immunodeficiency virus type 1 with zidovudine treatment, *New England Journal of Medicine*, **331**, 1173–1180. (Report written for the Pediatric AIDS Clinical Trial Group Protocol 076 Study Group).

Durham, S.D., Flournoy, N. and Li, W. (1998). A sequential design for maximizing the probability of a favorable response, *Canadian Journal of Statistics*, **26**, 479–495.

Eisele, J.R. (1994). The doubly adaptive biased coin design for sequential clinical trials. *Journal of Statistical Planning and Inference*, **38**, 249–262.

Faries, D.E., Tamura, R.N. and Andersen, J.S. (1995). Adaptive designs in clinical trials: Lilly experience (with discussions), *Biopharmaceutical Report*, **3**, 1–11.

Ferguson, T.S. (1967). *Mathematical Statistics: A Decision Theoretic Approach*, Academic Press, New York.

Field, C. and Smith, B. (1995). Robust estimation - a weighted maximum likelihood approach. *International Statistical Review*, **62**, 405–424.

Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. and Stahel, W.A. (1986). *Robust Statistics: The Approach Based on Influence Functions.* John Wiley and Sons, New York.

Iglewicz, B. (1983). Alternative designs: sequential, multi-stage, decision theory and adaptive designs. In *Cancer Clinical Trials: Methods and Practice*, eds. M.E. Buyse, J. Staquet and R.J. Sylvester, pp. 312–334. Oxford University Press.

Li, W. (1995). *Sequential Designs for Opposing Failure Functions.* Washington: College of Arts and Sciences, American University, doctoral dissertation.

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

ROUT, C.C., ROCKE, D.A., LEVIN, J., GOUWS, E., REDDY, D. (1993). A reevaluation of the role of crystalloid preload in the prevention of hypotension associated with spinal anesthesia for elective cesarean section, *Anesthesiology*, **79**, 262–269.

TAMURA, R.N., FARIES, D.E., ANDERSEN, J.S. AND HEILIGENSTEIN, J.H. (1994). A case study of an adaptive clinical trials in the treatment of out-patients with depressive disorder, *Journal of the American Statistical Association*, **89**, 768–776.

WARE, J.H. (1989). Investigating therapies of potentially great benefit: ECMO, *Statistical Science*, **4**, 298-340.

WEI, L.J. (1979). The generalized Polya's urn for sequential medical trials, *Annals of Statistics*, **7**, 291–296.

– – ––– (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** , 838–843.

YAO, Q. AND WEI, L.J. (1996). Play the winner for phase II/III clinical trials, *Statistics in Medicine*, **15**, 2413–2423.

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

ATANU BISWAS AND AYANEDRANATH BASU
APPLIED STATISTICS UNIT
INDIAN STATISTICAL INSTITUTE
203 B.T. ROAD
KOLKATA 700035, INDIA
E-mails: atanu@isical.ac.in
ayanbasu@isical.ac.in