

Adaptive Allocation Designs for Normal and Binary Treatment Responses

Madhumita Shome
Krishna Kumar Hindu Academy

Rahul Bhattacharya
Calcutta University

Abstract

A class of adaptive designs is formulated in two stages for clinical trials to favour the better performing treatment for further allocation in an efficient way. The first stage of the allocation consists in randomizing subjects to each treatment arm with equal probability and performing a test of equality of treatment effects. The resulting p value and the available estimate of a treatment difference measure is used to assign the incoming second stage subjects. Considering binary and normal responses, several exact and asymptotic properties of the proposed allocation are thoroughly examined and compared with the existing allocation designs.

Keywords: adaptive allocation, p-value, ethics.

1. Introduction

The primary objective of any clinical trial is to determine the efficacy of the competing treatments on the basis of the responses of the participants of the trial. Before the trial is actually conducted, the knowledge about the performance of the treatments is absent. In the absence of any information about treatment superiority, the usual practice is to consider each treatment equally important and assign equal number of subjects to each treatment arm. But such an allocation strategy is often criticised as the best and the worst treatments receive the same number of subjects. However, clinical trials involve human beings and therefore the ethical goal in such trials is to provide the best possible care for the subjects. Thus, adaptively designed allocation, (in particular, response adaptive allocation) is the natural choice for its ability to skew the assignment towards the treatment doing better using the available data. Two stage allocation design is one among the popular data dependent allocations, where the first stage data plays a crucial role in determining the sampling fraction of the second stage. The basic goal is naturally to assign a greater fraction of subjects to the treatment doing better. One of the simplest, perhaps the earliest, instance of two stage allocation in the field of a clinical trial can be found in [Colton \(1963\)](#), where for the assignment of n recruited subjects, $2m$ subjects, m on each treatment arm, are used in the first stage and the treatment doing better is selected for the assignment of the remaining $(n - 2m)$ subjects. In a further work, [Coad \(1992\)](#) studied the inferential properties of the procedure, such as exact bias and variance of the estimated treatment difference. Specifically, assuming that a higher response

is desirable, Coad considered the treatment producing the higher observed mean response in the first stage as better performing and suggested to assign the subjects of the second stage exclusively to this treatment. However, treatment assignments for the second stage subjects become perfectly predictable and hence lacks randomization (Atkinson and Biswas (2014), Antognini and Giovagnoli (2015)). Moreover, the best performing treatment after the first stage is decided through an examination of the sign of the estimated treatment difference and ignores the further information contained in the magnitude. Consequently, large as well as small positive values of the differences are given the same importance. Bandyopadhyay and Bhattacharya (2006) and Bandyopadhyay and Bhattacharya (2007) identified these issues and incorporating randomization in a convenient way, developed an ethical allocation after sequential determination of the first stage sampling allocation fraction. In further works, Bandyopadhyay, Biswas, and Bhattacharya (2009) and Bandyopadhyay, Biswas, and Bhattacharya (2010) used optimum design theory to develop ethical allocation for survival outcomes incorporating randomization in the second stage, where a prefixed allocation function is estimated on the basis of the first stage data and/ or incoming patients' covariate. However, in a recent work, Bhattacharya and Shome (2015) developed a two stage allocation procedure using the sufficient statistics based on the first stage data in a convenient way. To be specific, they used a decreasing function of the p value of a test of equality of treatment effects based on the first stage data to set the allocation probability of each subject of the second stage. The second stage allocation was fully randomized but the randomization probability remained the same. The present work develops an adaptive allocation procedure in two stages, where apart from being a function of the p value of a test of equality of treatment effects based on the first stage data, the second stage randomization probabilities are continuously updated after each response of the second stage assignments. The allocation procedure together with the related properties are discussed in Section 2 giving emphasis on the asymptotic properties. Considering binary and continuous treatment outcomes, the performance of the proposed procedure is compared with some relevant competitors in Section 3. Redesigning a real clinical trial adopting the proposed methodology is also added in Section 3. Finally Section 4 concludes with a discussion of a few relevant issues.

2. The allocation design

2.1. The general allocation function

Consider a clinical trial with two competing treatments (say, treatments A and B) and sequentially arriving subjects. A total of n subjects are to be assigned in two stages to accomplish certain objectives. The first stage assigns a total of $2m$ subjects to treatments A and B with equal probability (i.e. $\frac{1}{2}$) to get an initial idea about the treatment performances. We suggest to assign the $i + 1$ th subject of the second stage ($i \geq 0$) to treatment A with probability π_{mi} , where π_{mi} depends on the first stage data together with the allocation and response data of the i assigned subjects of the second stage. For a worthwhile determination of π_{mi} , we assume that a higher response is desirable and consider a statistic T_m based on the first stage data for an evaluation of the null hypothesis of equality of treatment effects with p_m as the corresponding p value. Also suppose that $\hat{\Delta}_{mi}, i \geq 1$ is a strongly consistent sequence of estimates of a treatment effect measure Δ (to be specified later), based on the first stage data and the allocation and response information available so far. Use of currently available estimate of treatment effect for the allocation of subjects was considered in Bandyopadhyay and Biswas (2001) with the goal of favouring the better performing treatment. But assigning more to the better performing treatment makes the allocation unbalanced and hence causes, in general, a significant loss in power. Therefore, as a trade off between favouring the better performing treatment and loss of statistical precision, we suggest to use the p value based on the first stage together with the available estimate of the treatment effect measure. Thus a sensible choice of π_{mi} should be a function of both p_m and $\hat{\Delta}_{mi}$, that is $\pi_{mi} = \phi(p_m, \hat{\Delta}_{mi})$ for

some function ϕ .

The most important task, is therefore, to suggest meaningful choices of ϕ following a thorough investigation of the properties. First of all, we note that the lower the value of p_m , the stronger is the evidence against H_0 and hence the more is the importance of the second stage data for further allocation. Thus it seems reasonable to assume π_{mi} as a non-increasing function of p_m for fixed $\hat{\Delta}_{mi}$. Again for any i , $\hat{\Delta}_{m,i+1} \geq \hat{\Delta}_{m,i}$ indicates higher superiority of treatment A at stage $i + 1$ than at stage i and hence an increased probability of assignment to treatment A at stage $i + 1$ than stage i is preferable. This naturally suggests that a desirable choice of ϕ must be non-decreasing in the estimates of the treatment effect measure for fixed p_m . However, treatment effect measure is, in general, unbounded and hence we use a reasonable monotonic and bounded function of the estimated treatment effect. Since we need to maintain a reasonable ordering in allocation probability, the most obvious choice of such a function becomes the distribution function of a random variable. Thus, we need a function $\phi(x, y) \in [0, 1]$ defined for $x \in [0, 1]$ and $y \in [0, 1]$ such that ϕ is non-increasing in x for fixed y and non-decreasing in y for fixed x . We provide below three such choices of ϕ considering different requirements.

For a meaningful development, assume that μ_k is the effect measuring parameter for treatment k , $k=A, B$ and consider testing $H_0 : \mu_A = \mu_B$ against the one sided alternative $H : \mu_A > \mu_B$. If we denote the p value related to testing H_0 against H by p_m , then an allocation probability for incoming second stage patient could be,

$$\phi^{(1)}(p_m, \hat{\Delta}_{m,i+1}) = 1 - p_m. \quad (1)$$

The above choice of ϕ is decreasing in p_m and provides the allocation design of [Bhattacharya and Shome \(2015\)](#), where the allocation probability is not updated after the first stage and hence is independent of any $\hat{\Delta}_{m,i}$.

With an aim to provide an allocation depending on both, the p value p_m and the sequence of updated estimates $\hat{\Delta}_{m,i}$, $i \geq 1$, we suggest the following allocation probability

$$\phi^{(2)}(p_m, \hat{\Delta}_{m,i+1}) = \left\{ G(\hat{\Delta}_{m,i}) \right\}^{w_{1m}}, \quad (2)$$

where G is a distribution function and w_{1m} is a function of p_m controlling the influence of the first stage outcome on the second stage allocation. It is easy to observe that a constant w_{1m} outweighs the importance of the first stage and hence a suitable function of p_m is appropriate. The choice (2) depends on $\hat{\Delta}_{mi}$ and meets all the necessary requirements provided w_{1m} is an increasing function of p_m . However, the author's choice for w_{1m} is $p_m + \frac{1}{2}$, a linear function of p_m with positive slope.

Finally, considering G as a distribution function and w_{2m} , a convenient function of p_m , we suggest the following second stage allocation probability

$$\phi^{(3)}(p_m, \hat{\Delta}_{m,i+1}) = \frac{1}{2}w_{2m} + \frac{1}{2}G(\hat{\Delta}_{m,i}). \quad (3)$$

Now for a meaningful determination of w_{2m} , consider testing $H_0 : \mu_A = \mu_B$ against the one sided alternative $H' : \mu_A < \mu_B$. If p_m is the p value associated to testing H_0 against H , defined earlier, then $1 - p_m$ can be regarded as the p value based on the same data related to testing H_0 against $H' : \mu_A < \mu_B$. Thus for a chosen significance level $100\alpha\%$, a value of p_m less than α indicates a strong support for $\mu_A > \mu_B$, whereas a value of $1 - p_m$ less than α gives a strong support in favour of $\mu_A < \mu_B$. Combining, we observe that the data gives strong support in favour of $\mu_A > \mu_B$ or $\mu_A < \mu_B$ as $p_m < \alpha$ or $p_m > 1 - \alpha$; otherwise the data gives evidence in favour of $\mu_A = \mu_B$, that is, equality of treatment effects. Therefore, for the assignment to treatment A, we must put a higher weight for $p_m < \alpha$ and a lower weight for $p_m > 1 - \alpha$. Moreover, for p_m lying between α and $1 - \alpha$, equal weights to both the

treatments are reasonable. This naturally suggests to use

$$\begin{aligned} w_{2m} &= 1 \text{ if } p_m < \alpha \\ &= \frac{1}{2} \text{ if } \alpha \leq p_m \leq 1 - \alpha \\ &= 0 \text{ if } p_m > 1 - \alpha. \end{aligned}$$

But such a choice does not always ensure unpredictability of treatment assignments and hence leads to selection bias (Rosenberger and Lachin (2002)). Thus considering the fact that the lower values of p_m and $1 - p_m$ indicate stronger evidence against the null hypothesis, we suggest the following

$$\begin{aligned} w_{2m} &= \frac{1}{2} \text{ if } \alpha \leq p_m \leq 1 - \alpha \\ &= 1 - p_m \text{ otherwise,} \end{aligned}$$

and use it extensively for further development.

2.2. Response driven choices of G and Δ

Now we come to the most crucial part of the development, that is, deciding an appropriate function G together with the form of treatment effect measure Δ . Although any distribution function is appropriate as G but such a choice induces arbitrariness. Consequently, we closely examine the meaning of $G(\hat{\Delta}_{m,i})$ and observe that $G(\hat{\Delta}_{m,i})$ is nothing but the estimated proportion of skewing to treatment A for the next (i.e. (i+1) th) subject, that is, a measure of the magnitude by which treatment A is better than treatment B. If X_k denotes the continuous outcome measure for treatment k , then the natural amount of skewing is simply $P(X_A > X_B)$ and hence an estimate can be used as $G(\hat{\Delta}_{m,i})$. For example, if the outcome measure for treatment k has a $N(\mu_k, \sigma_k^2)$ distribution, we have $P(X_A > X_B) = \Phi\left(\frac{\mu_A - \mu_B}{\sqrt{\sigma_A^2 + \sigma_B^2}}\right)$, which suggests to take $G = \Phi$, the cumulative distribution function of a standard normal variable and $\Delta = \frac{\mu_A - \mu_B}{\sqrt{\sigma_A^2 + \sigma_B^2}}$. As another example, we get $G(x) = \frac{x}{1+x}$, $x \geq 0$ and $\Delta = \frac{\mu_A}{\mu_B}$ when the response to treatment k is exponential with mean μ_k . It is interesting to note that if the treatment effect measure is a difference measure (e.g. normal responses), G is symmetric about the origin but for a ratio based treatment effect measure (e.g. exponential responses), the corresponding choice of G has a median at unity.

Motivated by the above discussion, one will be interested in either of $P(X_A > X_B)$ or $P(X_A \geq X_B)$ to use as the skewing proportion when X_k has a *Bernoulli*(μ_k) distribution. For independent Bernoulli responses X_A and X_B , it is easy to obtain $P(X_A > X_B) = \mu_A(1 - \mu_B)$ and $P(X_A \geq X_B) = 1 - \mu_B(1 - \mu_A)$. Now a close examination reveals that $P(X_A > X_B)$ and $P(X_A \geq X_B)$ can be both less than 50% even for $\mu_A > \mu_B$. Consequently, either $P(X_A > X_B)$ or $P(X_A \geq X_B)$ do not qualify as a reasonable skewing proportion. However, as an alternative, if we consider the average $\frac{1}{2}\{P(X_A > X_B) + P(X_A \geq X_B)\}$, then we get the simplified expression $\frac{1}{2} + \frac{1}{2}(\mu_A - \mu_B)$. Naturally, the simplified expression can be looked upon as $G(\Delta)$, where $\Delta = \mu_A - \mu_B$ and G is the distribution function of a uniform random variable over $[-1, 1]$. Then such an expression satisfies the requirements of a skewing proportion and hence can be used as a skewing proportion for binary responses. It is interesting to note that the resulting expression is nothing but the riddit function (Agresti (2010)) $P(X_A > X_B) + \frac{1}{2}P(X_A = X_B)$ based on binary responses. Thus, depending on the response distributions we get unique choices of G and the form of treatment effect measure Δ .

2.3. The allocation in practice

With the allocation indicator δ_i ($= 1$; if treatment A is assigned and $= 0$; otherwise) for the

i th subject, the allocation design can be described by the conditional probability

$$\begin{aligned} P(\delta_{i+1} = 1 | \mathcal{F}_i) &= \frac{1}{2} \text{ if } i \leq 2m - 1; \\ &= \phi(p_m, \hat{\Delta}_{m,i}) \text{ if } i \geq 2m, \end{aligned}$$

where \mathcal{F}_i is the information contained in the response and allocation data up to and including those of the i th patient, $i \geq 1$ with ϕ , p_m and $\hat{\Delta}_{m,i}$ are as defined before. If n assignments are made following the above allocation procedure, then the observed number of subjects to treatment A is simply $N_{An} = \sum_{i=1}^n \delta_i = n - N_{Bn}$, where δ_i are i.i.d Bernoulli variables with success probability $\frac{1}{2}$ for $i \leq 2m$. But exact properties of $\frac{N_{An}}{n}$ are not easy to express in tractable forms and therefore we proceed to examine the properties in large samples. For such an assessment, we assume $m = [n\theta]$, $0 < \theta < \frac{1}{2}$, where $[x]$ represents the greatest integer in x so that as $n \rightarrow \infty$,

$$m \rightarrow \infty \text{ but } \frac{m}{n} \rightarrow \theta. \quad (4)$$

The limiting value of the observed allocation proportion to treatment A, $\frac{N_{An}}{n}$ can be found in Result A.2 of the Appendix.

3. Exploring the performance measures

Since a reasonable allocation design aims to balance the ethical and statistical needs, it becomes important to examine both the ethical and inferential aspects. However, use of $P(X_A > X_B)$ or $P(X_A > X_B) + \frac{1}{2}P(X_A = X_B)$ as $G(\Delta)$ in the second stage is intended to assign more subjects to the better treatment. $G(\Delta)$ makes the allocation far from the balanced and hence a loss in statistical power is expected. On the other hand, equal randomization at the first stage reduces the power loss. Since, the proposed allocation designs are conducted in two stages comprising equal and adaptive randomization, a balance between ethics and statistical precision can be anticipated. For an assessment of the proposed designs from both the viewpoints(i.e. ethics and precision), we use the following measures

- The expected number of allocation to treatment A, $E(N_{An})$ (denoted by EN_{An}) along with its standard deviation and
- The power of a test of the equality of treatment effects and
- Error rate.

Considering normal and binary treatment responses, the above measures are computed for the proposed and competing allocation designs. In subsequent parts of this section, we refer the designs (1), (2) and (3) by D3, D1 and D2, respectively.

3.1. Normal responses

Suppose the response of patients to treatment k has a normal distribution with mean μ_k and variance σ_k^2 , $k = A, B$. Then the hypotheses of equality of treatment effects can be expressed as $H_0 : \mu_A = \mu_B$. If T_m is a statistic based on the first stage data relevant to the testing problem with T_m^{obs} as its observed value then p_m can be expressed as $P(T_m \geq T_m^{obs} | H_0)$. For normal responses with unknown and unequal variances, the natural choice of T_m is the Welch's t-statistic (Lehmann and Romano (2005)), defined by

$$T_m = \sqrt{m} \frac{\bar{X}_{Am} - \bar{X}_{Bm}}{\sqrt{s_{Am}^2 + s_{Bm}^2}}$$

where $\bar{X}_{km}(s_{km}^2)$ denotes the mean (variance) of the observed responses for the first stage patients assigned to treatment k . As indicated earlier, we have the choices $G = \Phi$ and $\hat{\Delta}_{m,i} = \frac{\hat{\mu}_{Ai} - \hat{\mu}_{Bi}}{\sqrt{\hat{\sigma}_{Ai}^2 + \hat{\sigma}_{Bi}^2}}, i \geq 2m$, where $\hat{\mu}_{ki}(\hat{\sigma}_{ki}^2)$ is the estimated value of $\mu_k(\sigma_k^2)$ after i responses are observed. For a meaningful comparative evaluation of the performance, we consider the following competitors.

• **A Two Stage Competitor:** As a two stage competitor, we consider the design of [Bhattacharya and Shome \(2015\)](#), where m subjects are assigned to each treatment arm in the first stage and based on these data, the p value p_m of a relevant test of treatment equality is calculated. Then each incoming subject of the second stage is assigned to treatment A with probability $1 - p_m$. Thus the procedure corresponds to the choice (1). Suppose N_{An} denotes the observed number of subjects assigned to treatment A out of n assignments. Then under (4), as $m \rightarrow \infty$, the observed allocation proportion to treatment A (i.e. $\frac{N_{An}}{n}$) approaches the almost sure limit $1 - \theta$ or θ according as $\mu_A - \mu_B > 0$ or $\mu_A - \mu_B < 0$ (see, Result A.2 of the Appendix and [Bhattacharya and Shome \(2015\)](#), for details). However, under $\mu_A = \mu_B$, p_m has a uniform distribution over $(0, 1)$ and hence in such a case $\frac{N_{An}}{n}$ converges in distribution to a uniform random variable, distributed over the interval $(\theta, 1 - \theta)$. Thus the limiting proportion of allocation depends on the sign of the treatment difference $\mu_A - \mu_B$ except for $\mu_A = \mu_B$.

• **Neyman Allocation:** We also consider a randomized version of Neyman allocation assuming (σ_A, σ_B) unknown. The allocation probability to treatment k for the $(i+1)$ th subject is $\frac{\hat{\sigma}_{ki}}{\hat{\sigma}_{Ai} + \hat{\sigma}_{Bi}}, k = A, B$, where $\hat{\sigma}_{ki}$ is the estimate (preferably, maximum likelihood estimate) of σ_k based on the first i allocation and response data ([Melfi, Page, and Gerald \(2001\)](#)). Use of sequentially updated maximum likelihood estimators of $\sigma_k, k = A, B$ results in a sequential maximum likelihood (SML) procedure ([Hu and Rosenberger \(2006\)](#)) targeting Neyman allocation. Now it is well known that under normality of responses, if $\sigma_k, k = A, B$ are known and the trial size is kept fixed (Chapter 2, [Rosenberger and Lachin \(2002\)](#)), such an allocation maximizes the power of Wald test for testing the equality of treatment effects. Thus Neyman allocation is *optimal* in such a situation. We, therefore, use this randomized version as a standard against which the inferential aspects of the designs can be assessed.

• **Bandyopadhyay and Biswas (2001) Allocation:** For a fair comparison, we also add the allocation of Bandyopadhyay and Biswas (2001)(referred to as BB) in the list of competitors. For such an allocation, the $(i + 1)$ th subject is allocated to treatment A with probability $\Phi(\frac{\hat{\mu}_{Ai} - \hat{\mu}_{Bi}}{T})$, where $T(> 0)$ is a tuning parameter and $\hat{\mu}_{ki}$ is the estimate of μ_k based on the first i allocation and response data. Naturally, the allocation probabilities are updated on the basis of the available estimates of the treatment effect.

Now for a valid assessment, we fix $n = 120$ and consider testing $H_0 : \mu_A = \mu_B$ against $H_1 : \mu_A > \mu_B$. Specifically, we take $\mu_A = \mu_B = 1$ under the null hypothesis and vary μ_A at regular intervals from 1.0 under the alternative. Naturally, treatment A is the better for the assumed configuration. For each configuration of (μ_A, μ_B) , we conduct a simulation study with 25,000 repetitions and compute expected number of allocations to the better treatment (i.e. treatment A) and power considering different choices of m and different combinations of (σ_A, σ_B) . The test statistic for the power computation is Welch's t statistic with unequal sample sizes, defined by

$$T_n^* = \frac{\bar{X}_{An} - \bar{X}_{Bn}}{\sqrt{\frac{s_{An}^2}{N_{An}} + \frac{s_{Bn}^2}{N_{Bn}}}},$$

where \bar{X}_{kn} and s_{kn}^2 are calculated on the basis of the responses from the two stages. Again, response adaptive randomization, in general, affects the type I error rate and such a rate depends on the convergence rate of the allocation proportions ([Yi and Wang \(2015\)](#)). Since, Welch's test is only approximate and asymptotically standard normal under the null hypothesis, we also report the attained significance levels (denoted by Error rate), where the nominal significance level is set at 5%. All these are provided in Table 1.

Table 1: Performance evaluation for normal responses with $n = 120$

Design D1	$EN_{AA}(SE)$ & Power		$EN_{AA}(SE)$ & Power		$EN_{AA}(SE)$ & Power	
	$\sigma_A = .25, \sigma_A = .25$		$\sigma_A = .5, \sigma_A = .25$		$\sigma_A = .25, \sigma_A = .5$	
$\mu_A - \mu_B$	$m = 15$	$m = 25$	$m = 15$	$m = 25$	$m = 15$	$m = 25$
0.0	60.54(14.42) .050 [.069]	60.38(11.26) .050 [.061]	60.03(14.54) .050 [.055]	60.13(11.26) .050 [.050]	61.01(14.38) .050 [.084]	60.66(11.21) .050 [.073]
0.06	71.52(13.30) .336	70.12(10.09) .356	67.11(14.03) .212	66.47(10.76) .209	68.02(13.96) .178	66.87(10.79) .184
0.12	80.90(10.60) .793	77.60(7.44) .810	73.92(12.62) .525	72.23(9.43) .524	74.64(12.75) .420	72.57(9.50) .435
0.18	87.69(7.75) .979	82.51(5.12) .984	79.81(10.75) .821	76.85(7.67) .821	80.37(11.10) .695	77.10(7.80) .715
0.24	92.30(5.66) .999	85.51(3.84) .999	84.60(8.89) .961	80.44(5.98) .963	85.08(9.35) .885	80.64(6.17) .901

Design D2	$EN_{AA}(SE)$ & Power		$EN_{AA}(SE)$ & Power		$EN_{AA}(SE)$ & Power	
	$\sigma_A = .25, \sigma_A = .25$		$\sigma_A = .5, \sigma_A = .25$		$\sigma_A = .25, \sigma_A = .5$	
0.0	59.99(12.10) .050 [.059]	60.01(9.58) .050 [.055]	59.79(12.19) .050 [.052]	59.94(9.60) .050 [.052]	60.13(12.18) .050 [.064]	60.11(9.60) .050 [.062]
0.06	68.06(12.96) .353	67.50(10.46) .348	64.80(12.44) .209	64.52(9.10) .206	65.26(12.75) .186	64.69(10.06) .190
0.12	76.77(13.55) .806	75.55(10.15) .811	70.07(13.02) .518	69.55(10.54) .520	70.54(13.48) .445	69.75(10.72) .457
0.18	84.93(11.96) .981	82.08(7.34) .984	75.50(13.31) .811	74.62(10.25) .813	75.99(13.76) .728	74.90(10.40) .743
0.24	91.23(8.82) .999	85.85(4.54) .999	80.88(12.79) .956	79.16(8.84) .959	81.29(13.27) .904	79.40(8.99) .919

Design D3	$EN_{AA}(SE)$ & Power		$EN_{AA}(SE)$ & Power		$EN_{AA}(SE)$ & Power	
	$\sigma_A = .25, \sigma_A = .25$		$\sigma_A = .5, \sigma_A = .25$		$\sigma_A = .25, \sigma_A = .5$	
$\mu_A - \mu_B$	$m = 15$	$m = 25$	$m = 15$	$m = 25$	$m = 15$	$m = 25$
0.0	59.93(26.32) .050 [.058]	60.03(20.50) .050 [.057]	59.86(26.31) .050 [.053]	59.92(20.46) .050 [.052]	59.99(26.29) .050 [.065]	60.01(20.45) .050 [.061]
0.06	75.91(24.15) .291	75.58(17.81) .320	70.25(25.44) .200	70.28(19.42) .221	70.25(25.40) .145	70.37(19.36) .163
0.12	88.87(18.60) .698	86.85(11.58) .756	79.73(22.82) .490	79.22(16.30) .532	79.80(22.85) .329	79.19(16.30) .367
0.18	97.31(12.23) .933	92.35(5.86) .966	87.73(19.23) .784	85.99(12.23) .821	87.68(19.33) .552	85.98(12.20) .615
0.24	101.92(6.82) .990	94.36(2.24) .998	93.96(15.04) .943	90.29(8.38) .961	93.94(15.11) .750	90.32(8.35) .818

BB	$EN_{AA}(SE)$ & Power		$EN_{AA}(SE)$ & Power		$EN_{AA}(SE)$ & Power	
	$\sigma_A = .25, \sigma_A = .25$		$\sigma_A = .5, \sigma_A = .25$		$\sigma_A = .25, \sigma_A = .5$	
$\mu_A - \mu_B$	$m = 2$	$m = 2$	$m = 2$	$m = 2$	$m = 2$	$m = 2$
0.0	60.01(6.04) .050 [.054]	60.01(6.04) .050 [.054]	59.77(7.00) .050 [.053]	59.77(7.00) .050 [.053]	60.24(7.00) .050 [.057]	60.24(7.00) .050 [.057]
0.06	62.75(6.07) .361	62.75(6.07) .361	62.53(6.96) .209	62.53(6.96) .209	63.04(7.05) .195	63.04(7.05) .195
0.12	65.46(6.03) .828	65.46(6.03) .828	65.21(6.93) .512	65.21(6.93) .512	65.72(7.05) .475	65.72(7.05) .475
0.18	68.13(6.01) .988	68.13(6.01) .988	67.87(6.85) .807	67.87(6.85) .807	68.41(7.09) .764	68.41(7.09) .764
0.24	70.84(5.94) .999	70.84(5.94) .999	70.59(6.77) .956	70.59(6.77) .956	71.13(7.01) .934	71.13(7.01) .934

Neyman	$EN_{AA}(SE)$ & Power		$EN_{AA}(SE)$ & Power		$EN_{AA}(SE)$ & Power	
	$\sigma_A = .25, \sigma_A = .25$		$\sigma_A = .5, \sigma_A = .25$		$\sigma_A = .25, \sigma_A = .5$	
$\mu_A - \mu_B$	$m = 2$	$m = 2$	$m = 2$	$m = 2$	$m = 2$	$m = 2$
0.0	59.98(8.89) .050 [.054]	59.98(8.89) .050 [.054]	80.14(8.33) .050 [.055]	80.14(8.33) .050 [.055]	39.86(8.32) .050 [.057]	39.86(8.32) .050 [.057]
0.06	60.02(8.74) .365	60.02(8.74) .365	80.11(8.24) .214	80.11(8.24) .214	39.27(8.27) .217	39.27(8.27) .217
0.12	59.94(8.79) .828	59.94(8.79) .828	80.08(8.28) .528	80.08(8.28) .528	39.90(8.29) .533	39.90(8.29) .533
0.18	60.03(8.88) .987	60.03(8.88) .987	80.11(8.28) .827	80.11(8.28) .827	39.90(8.31) .825	39.90(8.31) .825
0.24	60.02(8.80) .998	60.02(8.80) .998	80.08(8.27) .964	80.08(8.27) .964	39.92(8.22) .966	39.92(8.22) .966

Boldface figures indicate the powers for the proposed allocation and those within [] are the error rates.

Remarks. Before analyzing the numerical figures of Table 1, we provide a comparative picture of the second stage allocation probabilities for the two stage designs D1, D2 and D3. For the comparison, we fix $\mu_A = .25, \mu_B = 0, \alpha = .10$, consider $m = 25$ and $m = 30$ and vary (σ_A, σ_B) . For each configuration, second stage allocation probabilities (ϕ) are simulated for each allocation design at the sample sizes $n = 2m+1, 2m+2, \dots$ and plotted against the sample size n . All these can be found in figures 1 and 2. We observe that, $\phi^{(1)}$ remains fixed at a higher value, but $\phi^{(2)}$ and $\phi^{(3)}$ vary steadily at some lower values. This is expected, as $\phi^{(1)}$ does not take into account the second stage information and hence converges to a deterministic design. On the other hand, $\phi^{(2)}$ and $\phi^{(3)}$ consider the second stage information in addition to the first stage information and vary reasonably. However, for all the designs, the second stage allocation probabilities become stable, of course, with varying rates when sample size increases.

Now, we look at the performance measures, computed in Table 1. As indicated earlier, the performance assessment includes both the ethical (i.e. assigning a larger fraction to the better treatment) and inferential perspectives (i.e. detecting a departure in treatment effectiveness with high probability). First of all, consider the allocation designs D1, D2 and D3. For these designs, the $EN A_A$ figures increase with increase in $\mu_A - \mu_B$ irrespective of the choice of m and (σ_A, σ_B) . However, in any situation, higher number of patients are assigned to the better treatment (i.e. treatment A, in this case). But $EN A_A$ figures corresponding to $m = 25$ are, in general, lower than those with $m = 15$. This is natural because a large m causes most of the assignments through equal randomization and hence reduces the ethical allocation in the second stage. On the other hand, a lower m allows more allocation in the second stage, which in turn ensures more assignment to the treatment doing better.

Next, we consider the response adaptive competitors BB and Neyman allocations. Being purely response adaptive procedures, BB and Neyman allocation procedures do not depend on m . But to start the allocation, we require initial estimates and hence, we assign two subjects to each treatment arm and estimate the parameters and start adaptation from the fifth patient onwards. The allocation figures corresponding to BB are considerably lower than those corresponding to D1, D2 and D3. But Neyman allocation assigns more subjects to the treatment with higher response variability without caring for the treatment effectiveness and the expected allocation numbers remain stable at a value close to $n \left(\frac{\sigma_A}{\sigma_A + \sigma_B} \right)$. Moreover, apart from minor exceptions, Neyman allocation provides the highest statistical power. It is well known that for known but unequal response variabilities, Neyman allocation minimizes the trial size for fixed statistical precision (i.e. power), and the power figures of Table 1 are in well agreement with this even for unknown variabilities. However, BB allocation is seen to be less sensitive to small departure in treatment effectiveness and hence provides lower number of allocations together with higher statistical power as compared to D1, D2 and D3. This is consistent with the fact that unbalanced allocation, in general, causes a loss in power. Thus the proposed allocations (i.e. D1 and D2) assign a larger number of subjects to the better treatment with a little loss in statistical power and hence meets the specifications of a sensible allocation design.

Again from the figures of Table 1, we find that the attained significance level for all the allocation designs is slightly higher than the nominal level of 5%. Apart from minor exceptions, the difference from the nominal level is higher for smaller m than those associated with the higher choices of m . Therefore, adaptive allocation, in general, maintains a slightly higher significance level than the nominal one for moderate sample sizes.

Now, we shall investigate the behaviour of power for each allocation design with the increase in sample size. We set $\theta = \frac{1}{3}, \frac{1}{4}$ and compute the powers for testing $H_0 : \mu_A = \mu_B = 1$ against $H_1 : \mu_A \geq 1, \mu_B = 1$ for each allocation design. However, for brevity, we provide plots of power considering different combinations of (σ_A, σ_B) . The relevant plots are found

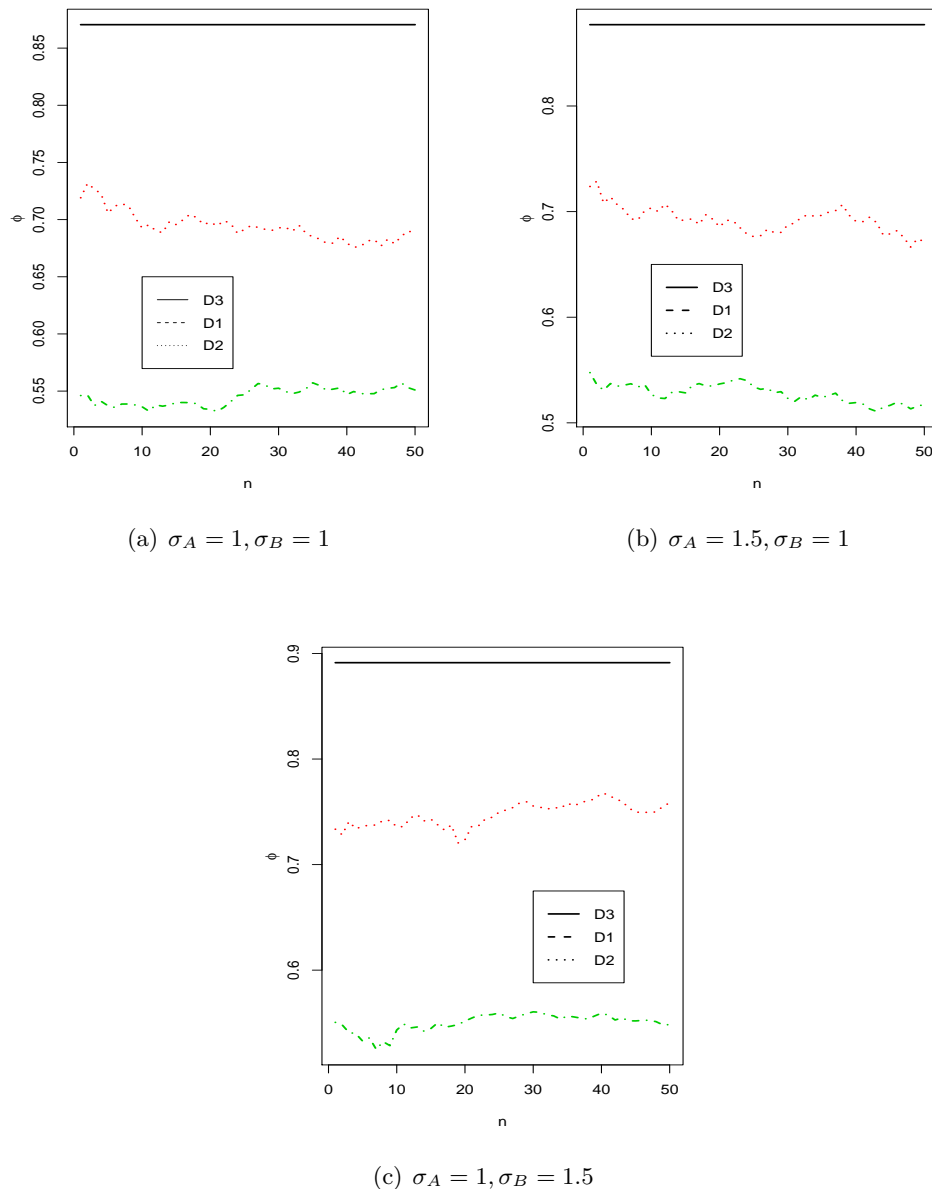


Figure 1: Second stage randomization probabilities for $m = 25$, $\mu_A = .25$, $\mu_B = 0.0$

in figures 3 and 4, where the Neyman allocation is indicated by O . It is easy to observe that, for all the allocation designs, the power figures increase with the sample size. However, rate of increase is not same for all the allocation designs. Specifically, if we consider Figure 3, we find that convergence rates of power are more or less same for $\sigma_A \geq \sigma_B$. But for $\sigma_A < \sigma_B$, D2 provides the highest and D3 provides the lowest figures. The powers for D3 is also the lowest for $\sigma_A = \sigma_B$. This is natural as D3 provides more skewing towards the better treatment and suffers a significant loss in power. Figure 4 mimics the same phenomena except that the power figures are lower than those for $\theta = \frac{1}{4}$, which is expected as a lower θ implies more allocation to the better treatment and hence results in a loss of power.

After considering the comparative picture discussed above, it is natural to set some recommendations for the practitioner. For all the proposed allocation designs and the competitors, the performance varies considerably with the configuration of (σ_A, σ_B) . In particular, the performance measures are encouraging for $\sigma_A \geq \sigma_B$ with respect to all the criteria. However, if we consider all the combinations of (σ_A, σ_B) , the allocation design D2 is seen to produce

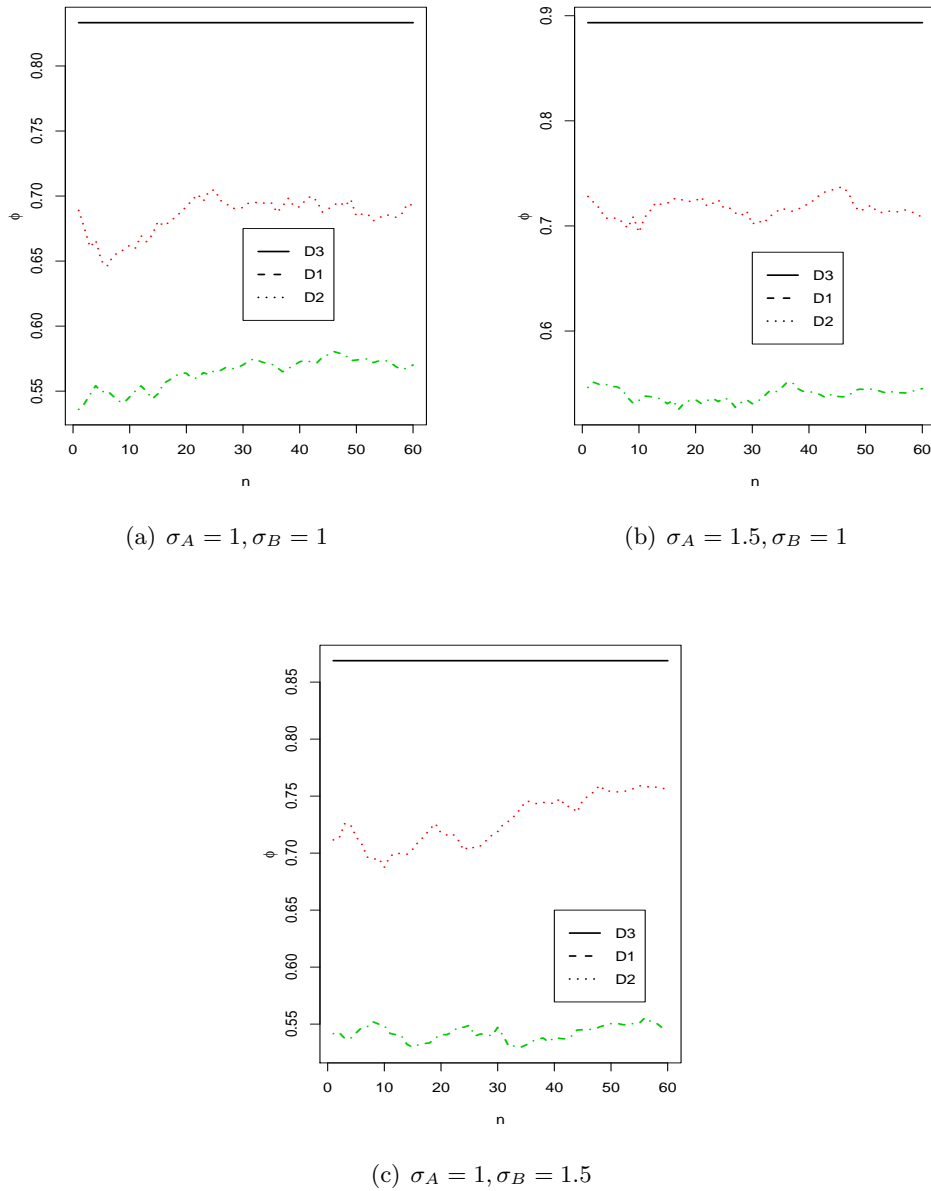
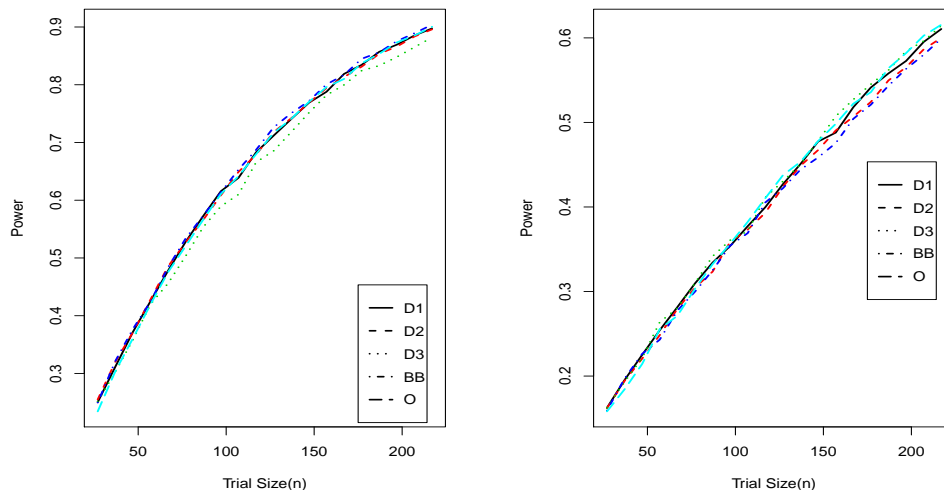


Figure 2: Second stage randomization probabilities for $m = 30$, $\mu_A = .25$, $\mu_B = 0.0$

promising results consistently. This naturally compels one to suggest the allocation design D2 for practical purposes.

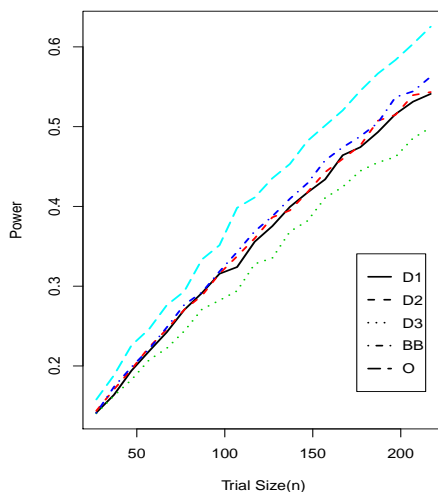
3.2. Binary responses

Suppose the response variable for treatment k is binary with success probability μ_k so that the hypotheses of equality of treatment effects is simply $H_0 : \mu_A = \mu_B$. We consider the estimated difference measure $T_m = \sqrt{m}(\frac{S_{Am}}{m} - \frac{S_{Bm}}{m})$ as the relevant statistic, where S_{km} is the number of successes by treatment k in the first stage. Now, for continuous responses, $P(T_m \geq T_m^{obs} | H_0)$ and $P(T_m > T_m^{obs} | H_0)$ are the same and hence either of these are used as the p value. But for discrete distributions, the difference between the two is consequential and hence as a compromise, we use the average of these two, called the mid-p value (see, [Agresti \(2013\)](#), for details), for our purpose. That is, we suggest to use $\frac{1}{2}P(T_m \geq T_m^{obs} | H_0) + \frac{1}{2}P(T_m > T_m^{obs} | H_0)$, the mid p value based on the first stage response and allocation data. Now under the null hypothesis, S_{Am} and S_{Bm} are independently and identically distributed as $Binomial(m, \mu)$,



(a) $\sigma_A = .25, \sigma_B = .25$

(b) $\sigma_A = .50, \sigma_B = .25$



(c) $\sigma_A = .25, \sigma_B = .50$

Figure 3: Comparing powers for testing $H_0 : \mu_A = \mu_B = 1$ against $H_1 : \mu_A \geq 1, \mu_B = 1$ with $\theta = \frac{1}{3}$.

where μ is the common but unspecified success probability under the null hypothesis. Writing $z = \sqrt{m}T_m^{obs}$, we find that $P(T_m \geq T_m^{obs}|H_0)$ is the same as $a_z = P(S_{Am} - S_{Bm} \geq z|H_0)$ with z as an integer in $\{-m, m\}$. Now a simple manipulation yields

$$\begin{aligned}
 a_z &= P(S_{Am} - S_{Bm} \geq z|H_0) \\
 &= P(S_{Am} \geq S_{Bm} + z|H_0) \\
 &= \sum_{0 \leq y, y+z \leq m} P(S_{Am} \geq y+z|H_0)P(S_{Bm} = y|H_0) \\
 &= \sum_{y=\max(0, -z)}^{\min(m, m-z)} P(S_{Am} \geq y+z|H_0)P(S_{Bm} = y|H_0).
 \end{aligned}$$

Since $P(T_m > T_m^{obs}|H_0)$ is the same as $P(S_{Am} - S_{Bm} \geq z+1|H_0)$, we get the equivalent expression $\frac{1}{2}a_z + \frac{1}{2}a_{z+1}$ for the mid p value. Naturally the above quantity depends on unknown

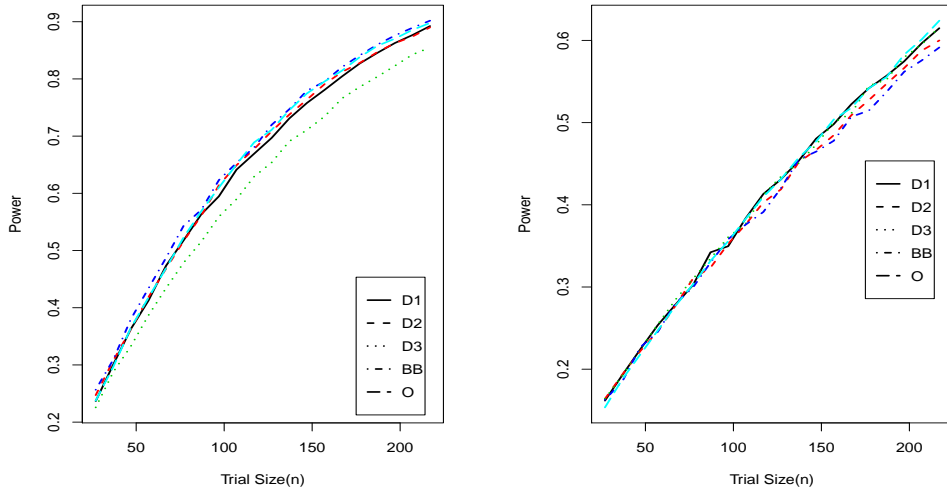
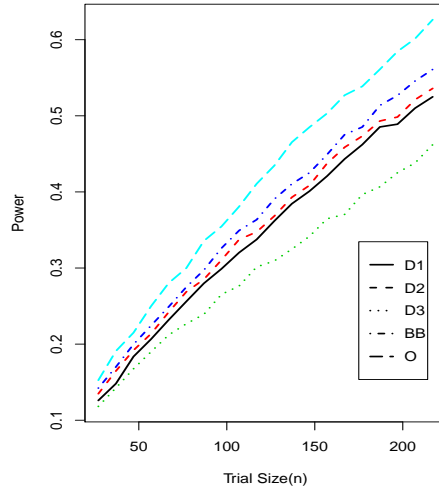
(a) $\sigma_A = .25, \sigma_B = .25$ (b) $\sigma_A = .50, \sigma_B = .25$ (c) $\sigma_A = .25, \sigma_B = .50$

Figure 4: Comparing powers for testing $H_0 : \mu_A = \mu_B = 1$ against $H_1 : \mu_A \geq 1, \mu_B = 1$ with $\theta = \frac{1}{4}$.

μ and hence we suggest to replace such a μ by its maximum likelihood estimate, that is, the pooled success proportion of the first stage. We use the estimated value of $\frac{1}{2}a_z + \frac{1}{2}a_{z+1}$ as our p_m for binary responses.

Now we consider the following competitors for a relative assessment of the proposed allocation designs.

- **Neyman Allocation:** As in Section 3.1, we consider a randomized version of Neyman allocation for binary responses, where the $(i+1)$ th subject is assigned to treatment k with probability $\frac{\sqrt{\hat{\mu}_{ki}(1-\hat{\mu}_{ki})}}{\sqrt{\hat{\mu}_{Ai}(1-\hat{\mu}_{Ai})} + \sqrt{\hat{\mu}_{Bi}(1-\hat{\mu}_{Bi})}}$, $k = A, B$, where $\hat{\mu}_{ki}$ is the maximum likelihood estimate of μ_k based on the first i allocation and response data.

• **Randomised-play-the-winner (RPW) Allocation:** RPW is a randomized allocation procedure for binary responses, developed by [Wei and Durham \(1978\)](#). The allocation uses an urn containing two types of balls representing two treatments. For the allocation of an incoming subject, a ball is drawn with replacement and the corresponding treatment is assigned. If the response is a success, an additional β balls of the same type are added to the urn; and for a failure, β balls of the opposite type are added. If the urn initially contains α balls of each type, the procedure is termed $RPW(\alpha, \beta)$. The rationale behind the allocation is to increase the chance of assigning the treatment *doing better*.

For the computation of the performance measures, we consider $n = 120$ and the allocation designs D1, D2, RPW and Neyman. Specifically, considering treatment A as the best treatment, we conduct a simulation study with 25000 repetitions for different m and calculate the relevant measures. As indicated earlier, the unspecified common success probability is estimated by the pooled success proportion, based on the first stage data. For the calculation of power, we use a statistic similar to T_m except that m is replaced by n and S_{km} by S_{kn} . All these are reported in Table 2.

Table 2: Performance evaluation for binary responses with $n = 120$

Design D1 $\mu_A - \mu_B$	$EN_A(SE)$ & Power for $\mu_B = .3$		$EN_A(SE)$ & Power for $\mu_B = .5$		$EN_A(SE)$ & Power for $\mu_B = .6$	
	$m = 20$	$m = 30$	$m = 20$	$m = 30$	$m = 20$	$m = 30$
0.0	63(8.52) .050[.058]	62.16(6.12) .050[.055]	62.12(12.12) .050[.057]	61.92(9.24) .050[.053]	63.36(12.00) .050[.056]	61.60(8.88) .050[.053]
0.1	71.92(6.36) .306	69.76(4.44) .319	71.04(10.8) .279	68.64(8.04) .283	71.76(12.00) .283	68.92(9.12) .273
0.2	78.4(4.8) .725	74.04(3.48) .751	77.88(8.52) .690	74.16(6.00) .702	78.28(10.44) .710	74.88(7.92) .704
0.3	82.16(3.96) .952	77.96(3.12) .965	82.80(6.12) .948	77.64(4.32) .958	83.10(7.92) .962	77.28(5.76) .965
Design D2 $\mu_A - \mu_B$	$EN_A(SE)$ & Power for $\mu_B = .3$		$EN_A(SE)$ & Power for $\mu_B = .5$		$EN_A(SE)$ & Power for $\mu_B = .6$	
	$m = 20$	$m = 30$	$m = 20$	$m = 30$	$m = 20$	$m = 30$
0.0	61.68(7.56) .050[.054]	60.04(5.20) .050[.052]	61.40(7.96) .050[.056]	60.84(6.42) .050[.053]	61.42(8.86) .050[.058]	60.81(6.49) .050[.054]
0.1	65.69(10.36) .333	65.52(8.06) .315	66.53(10.46) .279	65.39(8.17) .293	66.53(10.36) .258	65.32(8.08) .282
0.2	72.97(11.72) .759	71.02(9.57) .736	72.90(11.74) .685	71.08(9.52) .705	72.94(11.49) .692	71.10(9.38) .715
0.3	79.35(11.05) .966	76.83(7.07) .960	79.43(11.09) .948	76.78(7.09) .956	80.62(10.81) .969	77.87(6.98) .976
RPW(1,1) $\mu_A - \mu_B$	$EN_A(SE)$ & Power for $\mu_B = .3$		$EN_A(SE)$ & Power for $\mu_B = .5$		$EN_A(SE)$ & Power for $\mu_B = .6$	
	$m = 2$	$m = 2$	$m = 2$	$m = 2$	$m = 2$	$m = 2$
0.0	60.00(5.82) .050[.054]	60.00(5.82) .050[.054]	60.00(8.58) .050[.049]	60.00(8.58) .050.049	60.00(10.55) .050[.048]	60.00(10.55) .050[.048]
0.1	65.28(6.41) .308	65.28(6.41) .308	66.77(9.53) .287	66.77(9.53) .287	68.08(11.80) .289	68.08(11.80) .289
0.2	70.21(6.88) .721	70.21(6.88) .721	73.73(10.35) .705	73.73(10.35) .705	76.65(13.02) .724	76.65(13.02) .724
0.3	75.83(7.54) .952	75.83(7.54) .952	82.22(11.17) .953	82.22(11.17) .953	87.18(13.90) .963	87.18(13.90) .963
Neyman $\mu_A - \mu_B$	$EN_A(SE)$ & Power for $\mu_B = .3$		$EN_A(SE)$ & Power for $\mu_B = .5$		$EN_A(SE)$ & Power for $\mu_B = .6$	
	$m = 2$	$m = 2$	$m = 2$	$m = 2$	$m = 2$	$m = 2$
0.0	60.11(9.99) .050[.065]	60.11(9.99) .050[.065]	60.02(6.53) .050[.052]	60.02(6.53) .050[.052]	60.95(7.55) .050[.048]	61.05(7.55) .050[.048]
0.1	63.23(8.81) .289	63.23(8.81) .289	59.67(7.00) .285	59.67(7.00) .285	56.81(8.76) .298	56.81(8.76) .298
0.2	65.21(6.88) .700	65.21(6.88) .700	56.81(10.35) .710	56.81(10.35) .710	54.65(13.02) .742	54.65(13.02) .742
0.3	63.13(8.82) .950	63.13(8.82) .950	49.56(10.39) .963	49.56(10.39) .963	43.51(10.93) .981	43.51(10.93) .981

Boldface figures indicate the powers for the proposed allocation and those within [] are the error rates.

Remarks. From the performance measuring figures of Table 2, it is easily observed that the

proposed allocations maintain more or less a similar level of power with the Neyman allocation. Moreover, the number of allocations to the better treatment is always more than 50% and increases with increasing effectiveness of the better treatment. However, the performance of the proposed allocation designs (i.e. $D1$ and $D2$) remain similar for varying choices of μ_B . It is interesting to note that the allocation figures for $D1$ are mostly higher than those for the $RPW(1,1)$ allocation but both the procedures maintain a similar level of power. Again, as observed and explained for continuous responses, the number of allocation decreases with increase in m , the first stage allocation number. Thus, the proposed allocation designs assign more subjects to the better treatment keeping similar power level and hence is attractive from both the ethical and inferential viewpoints.

3.3. Redesigning a real trial

We further investigate the usefulness of the proposed designs considering the results of the real clinical trial on the treatment of patients with depressive disorder, conducted by Tamura, Faries, Andersen, and Heiligenstein (1994). The objective was to evaluate the clinical efficacy of *Fluoxetine* in the treatment of depressive disorder. First of all, we consider the data on patients with shortened rapid eye movement latency and consider the negative of the change in the first 17 items on the Hamilton Depression Scale after a minimum of 3 weeks therapy as the response. Then the response is continuous and a higher response indicates a favourable situation. Thus we have the response information on 42 patients, of which 21 were assigned to Fluoxetine and the rest to Placebo. For a valid assessment of the data, we prepare separate normal QQ plots for the response data on each treatment. The plots can be found in Figure 3 below.

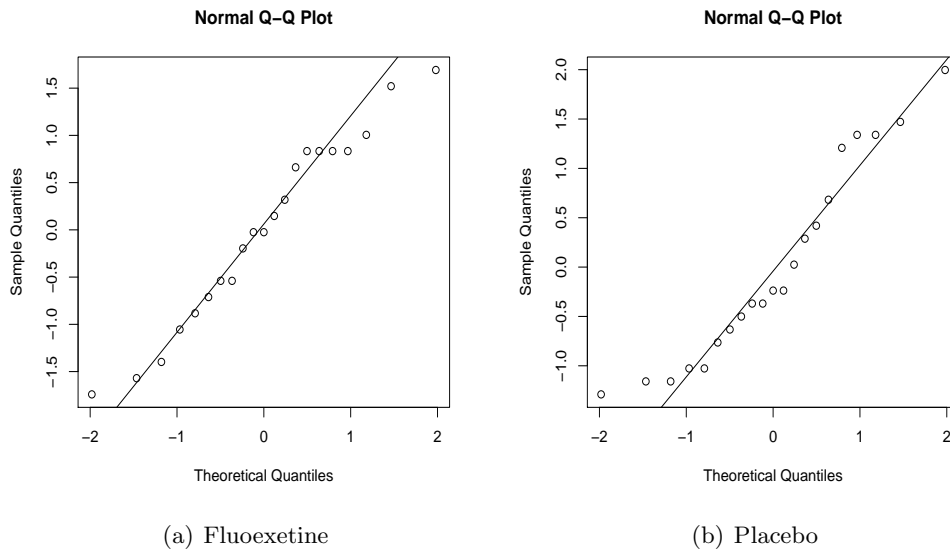


Figure 5: Normal QQ plots for the response on each treatment.

It is easy to observe that the QQ plots strongly support the normality of the underlying distributions. Assuming normality, we also conducted a test of equality of treatment effects and found a strong support (p value 1%) in favour of the superiority of Fluoxetine. We use these data as our first stage, where $m = 21$ assignments are made to each treatment and calculate mean responses and standard deviations as (11.14, 5.825) for Fluoxetine and (5.810, 7.607) for Placebo. Treating these summary measures as the true parameters for the respective normal response distributions, we carry out a simulation study considering different allocation designs and calculate the expected allocation proportion (EAP) to the

better treatment (i.e. Fluoxetine) for varying trial size(n). However, for brevity, we provide only a plot (see Figure 6) of the the EAP to Fluoxetine for different values of n adopting different allocation schemes.

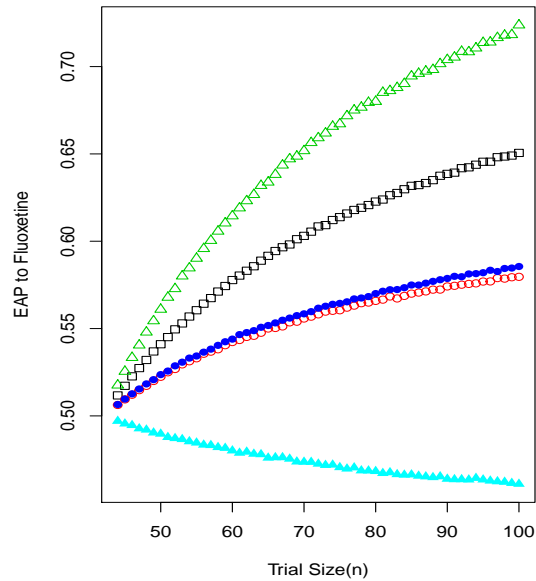


Figure 6: Comparing EAP to Fluoxetine in the redesigned trial considering normal responses for the allocation designs $D1(\square)$, $D2(\circ)$, $D3(\triangle)$, $BB(\bullet)$ and $Neyman(\blacktriangle)$.

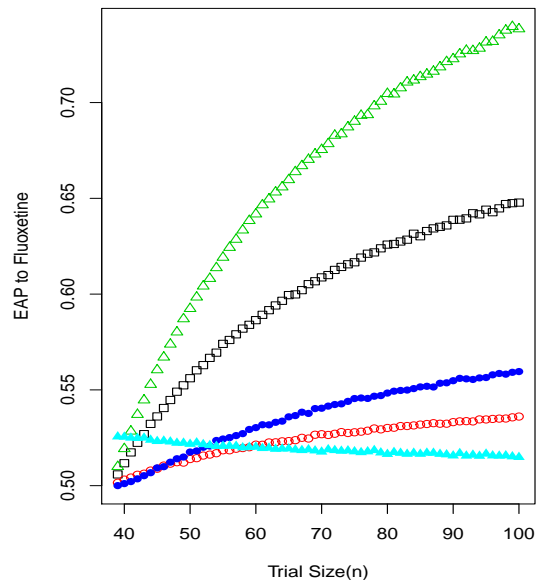


Figure 7: Comparing EAP to Fluoxetine in the redesigned trial considering binary responses for the allocation designs $D1(\square)$, $D2(\circ)$, $D3(\triangle)$, $RPW(\bullet)$ and $Neyman(\blacktriangle)$.

The nature of EAP for varying n indicates the usefulness of the proposed data dependent allocations (i.e. $D1$ and $D2$) over the competitors. However, the EAP for $D3$ is always higher

than those corresponding to D1 and D2. This is quite natural, as the first stage data provides a strong evidence of superiority of Fluoxetine and being a purely two stage design, D3 relies solely on the first stage for the second stage allocation and hence, on average, assigns more subjects to Fluoxetine. But D1 and D2 take into account both the first stage and available second stage data for further allocations in the second stage and hence are a little conservative in favouring Fluoxetine (i.e. the better treatment). For the response adaptive competitors (i.e. BB and Neyman) we use the available data of 42 assignments to obtain the starting estimates and start adaptation from the 43 rd patient onwards. BB allocation maintains an allocation proportion close to D2 but EAP for the Neyman allocation decreases from 50% as the better treatment has lower variability.

However, the final response of the actual trial was binary and considering shortened REML group, we get 39 responses (excluding the misclassified observations) of which 19 were from Fluoxetine and the remaining 20 from the Placebo arm. The number of successes were 11 and 7, from Fluoxetine and Placebo, respectively and hence we estimate the corresponding success rates as $\frac{11}{19}$ and $\frac{7}{20}$. Considering, these as the first stage data with $m = 19$, we, as earlier, carry out a simulation study considering different allocation designs to obtain the EAP figures to Fluoxetine for different values of n . For the calculation of the mid p value, we use the pooled proportion of success $\frac{11+7}{20+19} = \frac{18}{39}$ as an estimate of μ . However, we provide only a plot (see Figure 7) of the EAP figures to Fluoxetine for different values of n . As in the continuous responses, the plot shows increasing nature of EAP for varying n . As earlier, we use the summary measures based on 39 assignments to obtain initial estimates for the response adaptive competitors (i.e. RPW(1,1) and Neyman) and start adaptation from the 40 th patient onwards. EAP for RPW(1,1) is observed to be higher than those for D3 but lower than D1 and D2. However, EAP for the Neyman allocation decreases initially and then becomes stable at the limiting proportion 51%. These indicate the advantage of data dependent allocation over the fixed allocation design and hence make the proposed allocations as practitioner's one of the choices.

4. Concluding remarks

Adaptive allocation designs for normal and binary responses, proposed in the current work, show promising results over the existing competitors. Combinations of equiprobability and adaptive randomization made the proposed designs beneficial from the viewpoints of both clinician and statistician. However, the only subjectivity remains in the selection of the first stage sample size m . A smaller choice outweighs the importance of the first stage whereas a higher choice reduces ethical allocation. Therefore, an optimum choice is required for m . Since, the proposed allocation is an attempt to assign most of the subjects to the better performing treatment, one can search for an optimal m , which maximizes the expected number of allocation to the better treatment for fixed trial size n . It is easy to observe that for fixed (μ_A, μ_B) with $\mu_A > \mu_B$, the better treatment is A and the maximum value of $EN A_A$ is $m + (n - 2m) = n - m$. Hence, $m = 0$ maximizes the ethical benefit but simultaneously causes a loss in power. Therefore, as a compromise, we suggest to maximize the expected number of allocations to the better treatment subject to a fixed precision measure. However, such a choice is not immediate and we intend to work further on this issue.

Again, the allocation can be easily modified for responses like exponential or Weibull. But, these distributions are often used as life time distributions in situations where, in addition, some kind of censoring is present. Naturally, the corresponding development requires a rigorous treatment. Although the present development is based on two treatments, presence of multiple treatments is also possible. However, the corresponding development is not straightforward and depends on a number of issues. We intend to explore these aspects in a further work.

Appendix

Result A.1. (i) If the response to treatment k is Bernoulli with success probability μ_k , p_m is the p value as defined in Section 3.2 and $I(\cdot)$ is the indicator function, then as $m \rightarrow \infty$,

$$p_m \rightarrow I(\mu_A < \mu_B)$$

almost surely under any $\xi = (\mu_A, \mu_B)$ with $\mu_A \neq \mu_B$, and

$$p_m \rightarrow U$$

in distribution under any ξ with $\mu_A = \mu_B$, where U has a uniform distribution over the interval $(0, 1)$.

Proof: (i) Define two independent statistics $T_{0m} \sim F_{0m}$ and $T_{1m} \sim F_{1m}$, where F_{0m} is the cumulative distribution function (cdf) of T_m under $\mu_A = \mu_B$ and F_{1m} is the cdf of T_m under any ξ . Then we can equivalently express p_m as

$$\begin{aligned} p_m &= \frac{1}{2}P(T_{0m} \geq T_{1m}|T_{1m}) + \frac{1}{2}P(T_{0m} > T_{1m}|T_{1m}) \\ &= 1 - \frac{1}{2}\{F_{0m}(T_{1m}) + F_{0m}(T_{1m} - 0)\}. \end{aligned}$$

Now, as $m \rightarrow \infty$, T_{0m} converges in distribution to a $N(0, 2\mu(1 - \mu))$ random variable, where μ is the common unspecified value under the null hypothesis. Since the limiting cdf of T_{0m} (normal, in this case) is continuous, it follows from Pólya theorem (Billingsley (1995)), that as $m \rightarrow \infty$,

$$F_{0m}(T_{1m}) - \Phi\left(\frac{T_{1m}}{\sqrt{2\mu(1 - \mu)}}\right) \rightarrow 0 \tag{A.1}$$

almost surely, where Φ is the cdf of a standard normal variate.

Since the first stage of the allocation randomizes the subjects with equal probability to different treatment arms, the resulting observations are independent and identically distributed and hence it follows from the strong law of large numbers (Chapter 6, Billingsley (1995)) that as $m \rightarrow \infty$

$$\frac{S_{km}}{m} \rightarrow \mu_k$$

almost surely for $k = A, B$ and consequently, for any ξ ,

$$\frac{T_{1m}}{\sqrt{m}} \rightarrow \mu_A - \mu_B$$

almost surely. Thus, almost surely, $T_{1m} \rightarrow \infty$ or $T_{1m} \rightarrow -\infty$ as $\mu_A > \mu_B$ or $\mu_A < \mu_B$. Then it follows from (A.1) that as $m \rightarrow \infty$, both $F_{0m}(T_{1m})$ and $F_{0m}(T_{1m} - 0)$ converge almost surely to 1 or 0 as $\mu_A > \mu_B$ or $\mu_A < \mu_B$. Hence, it follows that as $m \rightarrow \infty$, $p_m \rightarrow I(\mu_A < \mu_B)$ under $\mu_A \neq \mu_B$.

However, under $\mu_A = \mu_B$, T_{0m} and T_{1m} are identically distributed and hence $\frac{T_{1m}}{\sqrt{2\mu(1 - \mu)}}$ is asymptotically standard normal. Thus $F_{0m}(T_{1m})$ and $F_{0m}(T_{1m} - 0)$ both converge in distribution to the same uniform random variable. Since, U and $1 - U$ have the same distribution for U having a uniform distribution over the interval $(0, 1)$, the proof concludes. \square

Note: For exponential and normal responses, Bhattacharya and Shome (2015) established the probability convergence of p_m . However, following Result A.1, almost sure convergence of p_m is straightforward.

Result A.2. If (2.4) holds, then under any ξ as $n \rightarrow \infty$

$$\frac{N_{An}}{n} \rightarrow \theta + (1 - 2\theta)\phi^*(\Delta)$$

almost surely, for some $\phi^*(\Delta)$.

Proof: Define \mathcal{G}_j as the total information contained in the j response and allocation data of the second stage. Now consider the representation,

$$\frac{N_{An}}{n} = \frac{1}{n} \sum_{j=1}^{2m} \delta_j + \frac{1}{n} \sum_{j=2m+1}^n \{\delta_j - E(\delta_j | \mathcal{F}_{2m})\} + \frac{1}{n} \sum_{j=2m+1}^n E(\delta_j | \mathcal{F}_{2m}), \quad (\text{A.2})$$

where \mathcal{F}_{2m} is the total response and allocation information of the first stage. From the strong law of large numbers for the i.i.d sequence of random variables and (4), we get

$$\frac{1}{n} \sum_{j=1}^{2m} \delta_j \rightarrow \theta$$

almost surely, as $m \rightarrow \infty$.

Now, to establish the convergence for the second term on the right hand side of (A.2), we define $M_k = \sum_{j=2m+1}^k \frac{1}{j} \{\delta_j - E(\delta_j | \mathcal{F}_{2m})\}$ for $k > 2m$. Then $\{E(M_k | \mathcal{F}_{2m}), k > 2m\}$ is a zero mean martingale with respect to $\{\mathcal{G}_j, j \geq 2m+1\}$ and consequently, it is easy to observe from the martingale convergence theorem (Hall and Heyde (1980)) and the fact that

$$|\delta_j - E(\delta_j | \mathcal{F}_{2m})| \leq 1 \text{ for any } j,$$

$E(M_k | \mathcal{F}_{2m})$ converges almost surely to zero. Applying Kronecker's Lemma we finally get

$$\frac{1}{n} \sum_{j=2m+1}^n \{\delta_j - E(\delta_j | \mathcal{F}_{2m})\} \rightarrow 0,$$

almost surely as $m \rightarrow \infty$. Again

$$E(\delta_j | \mathcal{F}_{2m}) = E\{\phi(p_m, \hat{\Delta}_{m,j}) | \mathcal{F}_{2m}\},$$

where ϕ is a bounded function. Hence it follows from the dominated convergence theorem, strong consistency of $\hat{\Delta}_{m,j}$ and Result A.1 that as $n \rightarrow \infty$, $E\{\phi(p_m, \hat{\Delta}_{m,j}) | \mathcal{F}_{2m}\}$ converges almost surely to $\phi(I(\mu_A < \mu_B), \Delta)$ or $E\{\phi(U, \Delta)\}$ as $\mu_A \neq \mu_B$ or $\mu_A = \mu_B$, respectively. Hence the last term in the righthand side of (A.2) converges almost surely to $(1 - 2\theta)\phi^*(\Delta)$, where

$$\begin{aligned} \phi^*(\Delta) &= \phi(I(\mu_A < \mu_B), \Delta) \text{ for } \mu_A \neq \mu_B, \\ &= E\{\phi(U, \Delta)\} \text{ for } \mu_A = \mu_B. \end{aligned}$$

Combining all these, we finally get the desired result. □

Acknowledgements

The authors are grateful to the Editor, Associate Editor and the anonymous Reviewer for their valuable suggestions, which helped to improve the quality of the manuscript.

References

- Agresti A (2010). *Analysis of Ordinal Categorical Data*. John Wiley & Sons, Inc., Hoboken, New Jersey.
- Agresti A (2013). *Categorical Data Analysis*. John Wiley & Sons, Inc., Hoboken, New Jersey.

- Antognini AB, Giovagnoli A (2015). *Adaptive Designs for Sequential Treatment Allocation*. Chapman and Hall.
- Atkinson AC, Biswas A (2014). *Randomised Response-Adaptive Designs in Clinical Trials*. CRC Press, Boca Raton.
- Bandyopadhyay U, Bhattacharya R (2006). “An Optimal Sequential Procedure In Ethical Allocation.” *Journal of Statistical Planning and Inference*, **136**(2), 430–446.
- Bandyopadhyay U, Bhattacharya R (2007). “On an Ethical Cum Optimal Adaptive Allocation Design.” *Statistics*, **41**(6), 471–483.
- Bandyopadhyay U, Biswas A (2001). “Adaptive Designs For Normal Responses With Prognostic Factors.” *Biometrika*, **88**(2), 409–419.
- Bandyopadhyay U, Biswas A, Bhattacharya R (2009). “A Bayesian Adaptive Design for Two-Stage Clinical Trials with Survival Data.” *Life Time Data Analysis*, **15**, 468–492.
- Bandyopadhyay U, Biswas A, Bhattacharya R (2010). “A Covariate-Adjusted Adaptive Design for Two-Stage Clinical Trials with Survival Data.” *Statistica Neerlandica*, **64**(2), 202–226.
- Bhattacharya R, Shome M (2015). “A Randomized Two Stage Allocation for Continuous Response Clinical Trials.” *Statistical Methods & Applications*, **24**(3), 373–386.
- Billingsley P (1995). *Probability and Measure*. John Wiley & Sons, Inc., New Jersey. ISBN 978-1482203530.
- Coad DS (1992). “Some Results on Estimation for Two Stage Clinical Trials.” *Sequential Analysis*, **11**(4), 299–311.
- Colton T (1963). “A Model for Selecting One of Two Medical Treatments.” *Journal of American Statistical Association*, **58**(302), 388–400.
- Hall P, Heyde CC (1980). *Martingale Limit Theory and Its Application*. Academic Press, London. ISBN 978-1482203530.
- Hu F, Rosenberger WF (2006). *The Theory of Response-Adaptive Randomization in Clinical Trials*. Wiley Interscience, New York. ISBN 978-1482203530.
- Lehmann EL, Romano JP (2005). *Testing Statistical Hypotheses*. Springer, New York. ISBN 978-1482203530.
- Melfi VF, Page C, Gerald M (2001). “An Adaptive Randomized Design with Application to Estimation.” *Canadian Journal of Statistics*, **29**(1), 107–116.
- Rosenberger WF, Lachin JL (2002). *Randomisation in Clinical Trials: Theory and Practice*. Wiley, New York. ISBN 978-1482203530.
- Tamura RN, Faries DE, Andersen JS, Heiligenstein JH (1994). “A Case Study of an Adaptive Clinical Trial in the Treatment of Out-Patients with Depressive Disorder.” *Journal of the American Statistical Association*, **89**(427), 768–776.
- Wei LJ, Durham S (1978). “The Randomized Play-The-Winner Rule in Medical Trials.” *Journal of the American Statistical Association*, **73**(364), 838–843.
- Yi Y, Wang X (2015). “Statistical Inference Following Response-Adaptive Randomization.” In, *Modern Adaptive Randomized Clinical Trials, Edited by Oleksandr Sverdlov*, **73**, 838–843.

Affiliation:

Rahul Bhattacharya
Department of Statistics
Calcutta University
35 Ballygunge Circular Road, Kolkata-700019, India
E-mail: rahul_bhatty@yahoo.com